

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:28:00 ; Search time 44 Seconds

(without alignments)
457.120 Million cell updates/sec

Title: US-09-705-911-24
Perfect score: 1401

Sequence: 1 MTA¹PFVSI¹IP¹TFNA¹VT¹LQ¹.....AL¹R¹TR¹L¹IR¹VK¹AV¹SK¹ER¹SA¹EP¹ 266

Scoring table: BLOSUM62

arched: 288829 seqs, 75613885 residues

Total number of hits satisfying chosen parameters: 288829

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published_Applications_AA:

- 1: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB pep: *
- 2: /cgn2_6/ptodata/2/pubppa/PTCT_NEW_PUB pep: *
- 3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB pep: *
- 4: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB pep: *
- 5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB pep: *
- 6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB pep: *
- 7: /cgn2_6/ptodata/2/pubppa/PTCTUS_PUBCOMB pep: *
- 8: /cgn2_6/ptodata/2/pubppa/PTCTUS_PUBCOMB pep: *
- 9: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB pep: *
- 10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB pep: *
- 11: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB pep: *
- 12: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB pep: *
- 13: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB pep: *
- 14: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	DB	ID	Description
1	17.9	12.8	278	10	US-09-767-041-36	Sequence 36, App
2	17.8	12.7	332	10	US-09-767-041-22	Sequence 22, App
3	17.2	12.3	358	10	US-09-815-242-5714	Sequence 5714, App
4	17.2	12.3	573	10	US-09-815-242-12747	Sequence 12747, App
5	16.5	11.8	150	10	US-09-924-358-29	Sequence 29, App
6	16.5	11.8	251	9	US-09-738-628-4202	Sequence 4202, App
7	16.3	11.6	332	10	US-09-767-041-21	Sequence 21, App
8	16.2	11.6	274	9	US-09-738-626-3905	Sequence 21, App
9	15.9	11.4	120	10	US-09-767-041-51	Sequence 3905, App
10	15.8	11.3	348	12	US-10-007-267-3	Sequence 51, App
11	15.8	11.3	348	12	US-10-007-267-11	Sequence 3, App
12	15.7	11.2	706	10	US-09-815-242-4950	Sequence 11, App
13	15.7	11.2	715	10	US-09-815-242-4950	Sequence 4950, App
14	15.6	11.2	270	10	US-09-816-0284-39	Sequence 10511, App
15	15.3	11.1	120	10	US-09-767-041-52	Sequence 39, App
16	15.3	10.9	322	10	US-09-767-041-34	Sequence 34, App
17	15.2	10.8	322	10	US-09-767-041-35	Sequence 35, App
18	15.0	10.7	187	9	US-09-973-457-4	Sequence 4, App
19	15.0	10.7	187	10	US-09-815-028-7	Sequence 7, App

20	150	10.7	187	12	US-10-074-527-4	Sequence 4, Appl1
21	149.5	10.7	972	9	US-09-879-959-10	Sequence 10, Appl1
22	146.5	10.5	301	9	US-09-816-028A-27	Sequence 27, Appl1
23	145.5	10.4	345	9	US-09-738-626-1894	Sequence 3894, Ap
24	144.5	10.3	281	10	US-09-765-272-196	Sequence 196, Appl
25	144.5	10.3	297	10	US-09-816-028A-31	Sequence 31, Appl
26	141	10.1	303	10	US-09-816-028A-22	Sequence 29, Appl
27	134.5	9.6	674	10	US-09-765-272-200	Sequence 200, Appl
28	132	9.4	210	9	US-09-767-041-47	Sequence 47, Appl
29	131	9.4	243	9	US-09-738-626-4733	Sequence 4733, Ap
30	129	9.2	389	10	US-09-816-028A-34	Sequence 34, Appl
31	128	9.1	337	12	US-10-007-267-5	Sequence 5, Appl1
32	128	9.1	337	12	US-10-007-267-12	Sequence 12, Appl
33	123.5	8.8	418	10	US-09-816-028A-33	Sequence 33, Appl
34	108.5	7.7	270	10	US-09-765-272-198	Sequence 198, Appl
35	108.5	7.7	303	10	US-09-765-272-202	Sequence 202, Appl
36	107.5	7.7	272	10	US-09-934-859-12	Sequence 12, Appl
37	107.5	7.7	272	10	US-09-934-858-32	Sequence 32, Appl
38	105	7.5	313	10	US-09-900-038A-1	Sequence 1, Appl1
39	102	7.3	440	9	US-09-902-525-45	Sequence 46, Appl1
40	100	7.1	317	10	US-09-934-859-18	Sequence 18, Appl
41	100	7.1	317	10	US-09-934-868-38	Sequence 38, Appl
42	96.5	6.9	395	12	US-10-042-523-4	Sequence 4, Appl1
43	94.5	6.7	398	10	US-09-741-659-430	Sequence 430, Appl
44	92.5	6.6	256	10	US-09-925-201-883	Sequence 883, Appl
45	92.5	6.6	269	10	US-09-767-041-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
HE-00-767-041-36

Sequence 36, Application US/09767041

; Patent No. US20020055168A1
: GENERAL INFORMATION:

APPLICANT: Smith, Hilda
TITLE OF INVENTION:

FILE REFERENCE: 2183-4726

CURRENT FILING DATE: 2001-01-22

PRIOR FILING DATE: 1999-07-19

PRIOR FILING DATE: 1998-07-22

PRIOR FILING DATE: 1998-07-22

```

; SOFTWARE: PatentIn version 3.0
END IN NO SC

```

LENGTH: 278

ORGANISM: Streptococcus suis

```
NAME/KEY: misc_feature
```

US-09-767-041-36

Query Match

Matches 6

QY 2 TAF

Db 3 TIS

QY. 62 LVV

Db 62 IRY

110 LGD QY

Db 122 DRV

Db 122 DRVDASGHFLTAEP LPTN-----QAVLSGRNVCKLLLEADGHRFVVACNKLYKK 170

OY 157 ELFDGIGYNYR---RWADMDENIRC---FSNPALITRYMDVVISSEYNDMTGFSMRG 210
 Db 171 ELFE-----DREFEKGKIHDEYFYRLYLELEKVAIVECLYYVDRENSITSSM--- 222
 OY 211 TDKEF-----RKRLPMY 222
 Db 223 TDHREHCLLEFONERMDRY 241

RESULT 2

US-09-767-041-22
 ; Sequence 22, Application US/09767041
 ; Patent No. US2002005168A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Hilda
 ; TITLE OF INVENTION: STREPTOCOCCUS SUTS VACCINES AND DIAGNOSTIC TESTS
 ; FILE REFERENCE: 2183-4726
 ; CURRENT APPLICATION NUMBER: US/09/767,041
 ; CURRENT FILING DATE: 2001-01-22
 ; PRIOR APPLICATION NUMBER: PCT/NL99/00460
 ; PRIOR FILING DATE: 1999-07-19
 ; PRIOR APPLICATION NUMBER: EP98202465.5
 ; PRIOR FILING DATE: 1998-07-22
 ; PRIOR APPLICATION NUMBER: EP98202467.1
 ; PRIOR FILING DATE: 1998-07-22
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: Patent version 3.0
 ; SEQ ID NO 22
 ; LENGTH: 332
 ; TYPE: PRT
 ; ORGANISM: Streptococcus suis
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: CPS2K
 US-09-767-041-22

Query Match 12.7%; Score 178; DB 10; Length 332;
 Best Local Similarity 21.9%; Pred. No. 5.9e-11;
 Matches 68; Conservative 53; Mismatches 100; Indels 90; Gaps 12;

OY 7 STIITPFAAVTLQACLSIVGQTYREVYVAVDGSITRITDIANSFRELGSRLVMS 66
 Db 5 STIVTYVEQTLKSCINSVQTKHLEILLVNDGSDTNSBEICLAYAKK-DSRIRYRK 63
 OY 67 GPDGPPYAMNKGAVANGVWFLGADDTLY-----EPTTLAQAFAFGDHA 114
 Db 64 KENGSLSPARNYGISRAKGDYLAFLDSDDFIHSEFIQRLHEALIERENALVAVAGYDRVDA 123
 OY 115 ASHLVYGVVWSTKSRHAGPPDLDLLEFETNLC-----HOSI-----FYRRELFDG 161
 Db 124 SGHFTLAEPLEPTN-----QAVLSGRNVCKKLEADGHRFVAVANNKLYKKELFD- 171
 OY 162 IGVYLAHY---RYWADMDENIRC---FSNPALITRYMDVVISSEYNDMTGFSMRGDKER 215
 Db 172 ---FREFEKGKIHDEYFYRLYLELEKVAIVECLYYVDRENSITSSM---TDHRE 223
 OY 216 ---RKRLPMY-----FWVAGWETCRRLM---AF 237
 Db 224 HCLLEFONERMDRYESRGDKELLLECYSFIAFAVILFLGKYHMLSKOOKKLOTLEIRIY 283
 OY 238 LKDKENRRIAL 248
 Db 284 KOLKONKRIAL 294

RESULT 3

US-09-815-242-5714
 ; Sequence 5714, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlssen, Karl L.

APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5714
 LENGTH: 358
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 US-09-815-242-5714

Query Match 12.3%; Score 172; DB 10; Length 358;
 Best Local Similarity 33.1%; Pred. No. 2.9e-10;
 Matches 51; Conservative 25; Mismatches 66; Indels 12; Gaps 5;

OY 6 FSIIITPFAAVTLQACLSIVGQTY---REVYVAVDGSITRITDIANSFRELG---S 60
 Db 1 FSVITVYNSKXITELTNSLAKQDPKTEFEVYVVDGSDTQIQVEYRKMLNKVMS 60
 OY 61 RLTVHSGPDGPPYAMNKGAVANGVWFLGADDTLYEPTTLAQAFAFGDHAASHLY 120
 Db 61 QLEITNSG---GPKKPRNVALKQAQGEVLEFVDSDDYI-NKETLKDAAFIDEHRHSDVL- 115
 OY 121 GDVWSTKSRHAGPPDLDLLEFETNLC-HOSITY 154
 Db 116 --TKMGVNGRGVPOSMFKETADEVTLLNSRITY 147

RESULT 4

US-09-815-242-12474
 ; Sequence 12474, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlssen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12474
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12474

```

```

Query Match 12.3%; Score 172; DB 10; Length 573;
Best Local Similarity 33.1%; Pred. No. 5.5e-10;
Matches 51; Conservative 25; Mismatches 66; Indels 12; Gaps 5;

```

```

OY 6 FSIITPTFAAATLQACISIVGQY--REVEVIVDGGSTRTDIDANSRPELG---S 60
DB 4 FSVIPTVSEKYTEILNLSLAKODEPKTEFEVYVDDCSTOITQIVERYKRLMLKYS 63
OY 61 RLVHSGPDDGPDYDAMNRGVATGEMVFLGADDTLYEPTTLAQVAALFGDHAASHLY 120
DB 64 QLEFNSG---GPKPRNVALKQAESEFVLFVDSDDYI--NKEILKDAALFIDEHSDVLL- 118
OY 121 GCVVARSSTRAGPDDRLLEFETNLCHQSIFY 154
DB 119 --IKKGVNGRVQPMKETAPEVTLNLSRIITY 150

```

```

RESULT 5
US-09-924-358-29
; Sequence 29, Application US/09924358
; Patent No. US20020107376A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; APPLICANT: Macbeth, Kyle
; TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
; TITLE OF INVENTION: 58764,
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE FAMILY MEMBERS AND USES THEREFOR
; FILE REFERENCE: 38155-20034.00
; CURRENT APPLICATION NUMBER: US/09/924,358
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/229,300
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid
US-09-924-358-29

```

```

Query Match 11.8%; Score 165.5; DB 10; Length 150;
Best Local Similarity 35.0%; Pred. No. 4.3e-10;
Matches 55; Conservative 26; Mismatches 63; Indels 13; Gaps 8;
OY 7 SIIITPTFAAATLQACISIVGQY--TYREVEVIVDGGSTDRDLIDANSRPELGSLVYH 65
DB 1 SIIITPTFAAATLQACISIVGQY--TYREVEVIVDGGSTDRDLIDANSRPELGSLVYH 59
OY 66 SPPDDGPDYDAMNRGVATGEMVFLGADDTLYEPTTLAQVAALFGDHAASHLYGY 123
DB 60 LEENIGLAAARAGLKHATGVDYIAFLDADDEV--PDMLEKILLEKNGAD-ITYGRV 116
OY 124 VNRSTRHAGPDDRLLEFETNLCHQSIFYRRELE 160

```

```

DB 117 I-NENKGRINGKLRL--LVF---LIGSNALYRREAL 147

```

```

RESULT 6
US-09-738-626-4202
; Sequence 4202, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SETKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHITO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: patentin ver. 3.0
; SEQ ID NO 4202
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4202

```

```

Query Match 11.8%; Score 165; DB 9; Length 251;
Best Local Similarity 27.4%; Pred. No. 1e-09;
Matches 76; Conservative 43; Mismatches 100; Indels 58; Gaps 15;

```

```

OY 1 MTAAPV---FSIITPTFAAATLQACISIVGQYREVEVIVDGGSTDRDLIDANSRPE 57
DB 1 MTSIGRISIVPCANDAVLLRQCLSSIVAGTLQPEVIVDGGSTDRDINSVEVAN----R 56
OY 58 LGSRLVHSGPDDGPDYDAMNRGVATGEMVFLGADDTLYEPTTLAQVAALFGDHAASH 117
DB 57 MGAR-VVHE--PLOGITWASAGYNSANGDLIVRDA--DCVIPPDLHLSQVNAIYN--RTE 110
OY 118 LVYGDVVMSTRKSRHAGPDDRLLEFETNLCHQSIFYRRELEFGIGPYNLRYVMADWD 176
DB 111 ETEGRTVALTGT---GSFPIGRMGWALCTYIGA--YRNSTKALCHYPI----- 157
OY 177 ENIRCESNPALITRYM-----DVVISE---YNDM-TGFSMKOCTDKEFRKRLPMYF--- 223
DB 158 -----FGSNSVIRQWMDYKQDITLSETPVHEMDYSEFVRPEHYEWEKRLNLINHR 212
OY 224 -----WAGNETCRMAFLKDKENRLALR 249
DB 213 ALMGVRSQFIRLVGFYTIK--IAWKREPVHORELSR 247

```

```

RESULT 7
US-09-767-041-21
; Sequence 21, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460

```


OPERATING SYSTEM: PC-DOS/MS-DOS

PRIOR FILING DATE: 2000-03-26
PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR APPLICATION NUMBER: 60/242,578

RESULT 15

US-09-767-041-52

Sequence 52, Application US/09767041

Patent No. US20020055168A1

GENERAL INFORMATION:

APPLICANT: Smith, Hilda

TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS

FILE REFERENCE: 2183-4726

CURRENT APPLICATION NUMBER: US/09/767,041

PRIOR FILING DATE: 2001-01-22

PRIOR FILING DATE: 1999-07-19

PRIOR APPLICATION NUMBER: EP98202465.5

PRIOR FILING DATE: 1998-07-22

PRIOR APPLICATION NUMBER: EP98202467.1

NUMBER OF SEQ ID NOS: 53

SOFTWARE: PatentIn version 3.0

SEQ ID NO 52

LENGTH: 120

TYPE: PRT

ORGANISM: Streptococcus suis

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: N-terminal part of CPS2K

NAME/KEY: misc_feature

LOCATION: (1)..(120)

OTHER INFORMATION: Xaa may be any amino acid

US-09-767-041-52

Query Match

Best Local Similarity 11.1%; Score 155.5; DB 10; Length 120;

Matches 32; Conservative 23; Mismatches 36; Indels 1; Gaps 1;

OY 7 SIITPENAATVLOACIGSVGTYREVEVLVDGSGTDRDIDIANSPRELGSRPLVHS 66

DB 5 SIIVPIYNEOYLKSCINCINIVNQYKHIELVNDGSSITDENSEICLAIYAKK-DSRIIRYFK 63

OY 67 GPDGPGYDAMNRGVATGEWVFLGADDTLY 98

DB 64 KENGISDARNYGISRAKGDYLAFLIDSDDFIH 95

Search completed: April 17, 2003, 17:36:37
Job time : 46 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:25:41 ; Search time 29 Seconds

(Without alignments)
269.879 Million cell updates/sec

Title: US-09-705-911-24

Sequence: 1 MTAPVFSIIIPFNNAVTLQ.....ALKRLIRKAKSKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Marched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued_Patents_AA:*
1: /cgn2_6/pdata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/pdata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/pdata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/pdata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/pdata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	161	11.5	324	1	US-08-597-236-10
2	161	11.5	324	1	US-08-746-682A-10
3	158	11.3	348	1	US-08-312-387B-3
4	158	11.3	348	1	US-08-312-387B-11
5	158	11.3	348	1	US-08-683-426-3
6	158	11.3	348	1	US-08-683-426-11
7	158	11.3	348	1	US-08-683-426-3
8	158	11.3	348	1	US-08-683-426-11
9	158	11.3	348	1	US-08-683-426-3
10	158	11.3	348	2	US-08-878-360-3
11	158	11.3	348	2	US-08-878-360-11
12	158	11.3	348	3	US-08-478-140B-3
13	158	11.3	348	4	US-08-478-140B-8
14	158	11.3	348	4	US-09-333-412-3
15	158	11.3	348	4	US-09-333-412-11
16	158	11.3	348	4	US-09-338-943-3
17	147.5	10.5	965	4	US-09-338-943-8
18	144.5	10.3	281	4	US-09-437-277-3
19	141	10.1	702	4	US-08-961-083-196
20	135.5	9.7	93	4	US-08-858-207A-521
21	134.5	9.6	674	4	US-08-961-083-200
22	128	9.1	337	1	US-08-312-387B-5
23	128	9.1	337	1	US-08-312-387B-12
24	128	9.1	337	1	US-08-683-426-5
25	128	9.1	337	1	US-08-683-426-12
26	128	9.1	337	1	US-08-683-426-5
27	128	9.1	337	1	US-08-683-426-12

28	128	9.1	337	2	US-08-878-360-5	Sequence 5, Appl1
29	128	9.1	337	2	US-08-878-360-12	Sequence 12, Appl1
30	128	9.1	337	3	US-08-478-140B-5	Sequence 5, Appl1
31	128	9.1	337	4	US-09-333-412-5	Sequence 5, Appl1
32	128	9.1	337	4	US-09-333-412-12	Sequence 12, Appl1
33	128	9.1	337	4	US-09-338-943-5	Sequence 5, Appl1
34	115.5	8.2	418	4	US-09-134-001C-4051	Sequence 4051, Ap
35	108.5	7.7	270	4	US-08-961-083-198	Sequence 198, Ap
36	108.5	7.7	303	4	US-08-961-083-202	Sequence 202, Ap
37	99	7.1	358	4	US-09-134-001C-5633	Sequence 5633, Ap
38	96.5	6.9	395	4	US-08-635-552A-4	Sequence 4, Appl1
39	96.5	6.9	419	2	US-08-270-581-2	Sequence 2, Appl1
40	96.5	6.9	419	4	US-09-146-893-2	Sequence 2, Appl1
41	92	6.6	727	4	US-09-134-001C-4067	Sequence 4067, Ap
42	84.5	6.0	79	4	US-08-961-083-168	Sequence 168, Ap
43	76	5.4	723	4	US-08-960-048-9	Sequence 9, Appl1
44	76	5.4	723	5	PCT-US91-01726-4	Sequence 4, Appl1
45	76	5.4	3724	2	US-08-804-227C-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1
US-08-597-236-10
Sequence 10, Application US/08597236
Patent No. 5733765
GENERAL INFORMATION:
APPLICANT: STINGELE, Francesca
APPLICANT: MOLTET, Beat
TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
TITLE OF INVENTION: EXOPOLYSACCHARIDES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americans
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/597,236
FILING DATE:
CLASSIFICATION: 426
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95201669.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci A., Allan
REGISTRATION NUMBER: 30256
INFORMATION FOR SEO ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-597-236-10

Query Match 11.5%; Score 161; DB 1; Length 324;
Best Local Similarity 35.9%; Pred. No. 2e-10;
Matches 42; Conservative 17; Mismatches 44; Indels 14; Gaps 3;

QY 5 VFIIIPFNNAVTLQACISIVGQYREVENVLDGSDTRTDIANSPFELGSRVY 64
DB 6 LIIIVPVVVEKYLECLOSVOQNTNNREVIILVNDGSDSSLSICEKFNQ-DKRSV 64
QY 65 HSPDDGPPIDAMNKGAVANGEWTLFGADDTLEPTTLQAFAFLGDHAAASHLYVG 121

[illegible]

DB 123 MGAMLEVLSEERKGNRLAHKHKIKMKPTRHEDIAAF-----PFGNPINNTMIMRRS 178
QY 158 LFDGIGPYNLRYRWADW 175
DB 179 VIDG-----GLRYDTERDM 192

RESULT 9
US-08-878-360-3
Sequence 3, Application US/08878360
Patent No. 5945322

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 343-1684
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-3

Query Match 11.3%; Score 158; DB 2; Length 348;
Best Local Similarity 23.7%; Pred. No. 4.9e-10;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

QY 4 PVSIIIPFENAAVTLQACISIVGQTYREVEVVLVDGSGTDRDLTIANSPRELGSRLV 63
DB 3 PLVSVLICAVNVEKYFAOSLAIVNQTWRNLDLIVDGSSTGTALIAKDFQRDSRIKI 62
QY 64 VHSPPDDGYDAMRGV-----GVATGEWVLFAGD-----TLYEPTTLAQ 105
DB 63 LAQAQNSGLIPSLINIGDLAKSGGGGGGYIARTDADDIASPGWIKIYGEEMKDSITA 122
QY 106 VAAFL-----GDHAAHLVYGDVVMRSYKSRHAGFPDLRLLEFETNLCHOSIFYRRE 157
DB 123 MGAMLEVLSEERKGNRLAHKHKIKMKPTRHEDIAAF-----PFGNPINNTMIMRRS 178
QY 158 LFDGIGPYNLRYRWADW 175
DB 179 VIDG-----GLRYDTERDM 192

DB 179 VIDG-----GLRYDTERDM 192

RESULT 10
US-08-878-360-11
Sequence 11, Application US/08878360
Patent No. 5945322

GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 343-1684
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-11

Query Match 11.3%; Score 158; DB 2; Length 348;
Best Local Similarity 23.7%; Pred. No. 4.9e-10;
Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

QY 4 PVSIIIPFENAAVTLQACISIVGQTYREVEVVLVDGSGTDRDLTIANSPRELGSRLV 63
DB 3 PLVSVLICAVNVEKYFAOSLAIVNQTWRNLDLIVDGSSTGTALIAKDFQRDSRIKI 62
QY 64 VHSPPDDGYDAMRGV-----GVATGEWVLFAGD-----TLYEPTTLAQ 105
DB 63 LAQAQNSGLIPSLINIGDLAKSGGGGGGYIARTDADDIASPGWIKIYGEEMKDSITA 122
QY 106 VAAFL-----GDHAAHLVYGDVVMRSYKSRHAGFPDLRLLEFETNLCHOSIFYRRE 157
DB 123 MGAMLEVLSEERKGNRLAHKHKIKMKPTRHEDIAAF-----PFGNPINNTMIMRRS 178
QY 158 LFDGIGPYNLRYRWADW 175
DB 179 VIDG-----GLRYDTERDM 192

RESULT 11
US-08-478-140B-3

[illegible]

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/333,412
 FILING DATE: 15-Jun-1999
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/312,387
 FILING DATE: July 7, 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-095
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201 487-5800
 TELEFAX: 201 343-1684
 TELEX: 133521
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 348 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 11:
 US-09-333-412-11

Query Match 11.3%, Score 158; DB 4; Length 348;
 Best Local Similarity 23.7%; Pred. No. 4.9e-10;
 Matches 47; Conservative 37; Mismatches 80; Indels 34; Gaps 5;

QY	4	PVEFTIIITFNAAVYLQACISIVGQTYREVEVVLVDGGSTRTDIDIANSPREPGLSRIV	63
DB	3	PLVSALLCAVNEKYEFAOSLAAVNQMWNMDILLVDDSGTDGTILAIADFORSRKI	62
QY	64	VHSGDDPEPYAMNGV-----GVATGEVYLFEGAD-----TLYEPTTLAQ	105
DB	63	LQAQNNSLIPSLNTIGDELAKSGGCGGYIARTDADDIASPGWIEKIYGENEKDRSILA	122
QY	106	VAALF-----GDHAASHLVYGDVMKSTKSRRHAGPEDLDRLLETNMICQSIFYYRE	157
DB	123	MGAWELEVLSEEDGRNLRAHHKHGRIWKPKTHEDIAAF----FPFGNIHNHTMIMRRS	178
QY	158	LEDGIGPYNLRIRYVADM	175
DB	179	VIDG---GLRYDTERDW	192

RESULT 15
 US-09-338-943-3
 Sequence 3, Application US/09338943
 Patent No. 6379933
 GENERAL INFORMATION:
 APPLICANT: JOHNSON, KARL F.
 APPLICANT: ROTH, STEPHEN
 APPLICANT: BOCCALA, STEPHANIE L.
 TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
 TITLE OF INVENTION: SACCARIDE UNITS WITH A POLYGLYCOSYLTANSFERASE, A
 TITLE OF INVENTION: POLYGLYCOSYLTANSFERASE AND GENE ENCODING A
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/338,943
 FILING DATE:

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:23:35 ; Search time 33 Seconds

(without alignments)
1650.865 Million cell updates/sec

Title: US-09-705-911-24

Perfect score: 1401
Sequence: 1 MFAPVSIITIPFNAATLQ.....ALTRLIRKAVSKERSAEP 266

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Targeted: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*
15: SP-virus:*
16: SP-bacteriaph:*
17: SP-archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1401	100.0	266	2	088109 mycobacteri
2	298.5	21.3	297	2	093025 salmonella
3	298	21.3	297	2	093025 salmonella
4	276	19.7	265	2	088109 mycobacteri
5	275	19.6	265	2	088109 mycobacteri
6	275	19.6	265	2	088109 mycobacteri
7	275	19.6	265	2	088109 mycobacteri
8	274	19.6	265	2	088109 mycobacteri
9	271	19.3	265	2	088109 mycobacteri
10	267.5	19.1	248	2	088109 mycobacteri
11	267.5	19.1	248	2	088109 mycobacteri
12	266.5	19.0	248	2	088109 mycobacteri
13	255.5	18.2	251	2	0926K3 leptospira
14	255	18.2	254	2	0926K3 leptospira
15	254.5	18.2	263	2	0926K3 leptospira
16	252	18.0	254	2	0926K3 leptospira

17	248	17.7	254	2	09XDJ3	09XDJ3 bacterioides
18	247	17.6	276	2	09R094	09R094 rhizobium e
19	238.5	17.0	247	2	09R094	09R094 yersinia ps
20	238.5	17.0	247	16	09R094	09R094 yersinia ps
21	232.5	16.6	260	16	09ZMX1	09ZMX1 helicobacte
22	231.5	16.5	259	16	024928	024928 helicobacte
23	220	15.7	324	16	08YSL1	08YSL1 anabaena sp
24	207	14.8	321	16	08YSM2	08YSM2 anabaena sp
25	207	14.8	368	16	098JH2	098JH2 rhizobium l
26	206.5	14.7	316	16	08YSL6	08YSL6 anabaena sp
27	205.5	14.7	298	17	09U2I6	09U2I6 pyrococcus
28	204.5	14.6	330	16	08YSL1	08YSL1 anabaena sp
29	201.5	14.4	321	2	09ATH3	09ATH3 streptococc
30	200	14.3	314	2	09L8A8	09L8A8 aeromonas h
31	196.5	14.0	337	16	08YSL6	08YSL6 anabaena sp
32	196.5	14.0	367	16	092V61	092V61 rhizobium m
33	192	13.7	336	16	097H39	097H39 clostridium
34	190.5	13.6	323	16	08YSL7	08YSL7 anabaena sp
35	190	13.6	248	16	08YSL7	08YSL7 anabaena sp
36	189.5	13.5	333	16	08YSL6	08YSL6 anabaena sp
37	188.5	13.5	333	16	097H38	097H38 pseudomonas
38	188	13.4	299	2	P95448	P95448 pseudomonas
39	188	13.4	299	16	09J5M1	09J5M1 pseudomonas
40	186.5	13.3	281	2	09ZGK2	09ZGK2 leptospira
41	186.5	13.3	316	2	093DZ7	093DZ7 streptococc
42	186	13.3	277	2	086296	086296 mycobacteri
43	186	13.3	322	2	09ALX0	09ALX0 streptococc
44	184	13.1	411	17	08YSL3	08YSL3 methanocarc
45	183.5	13.1	281	2	09S4F8	09S4F8 leptospira

ALIGNMENTS

RESULT 1

ID 088109 PRELIMINARY; PRT; 266 AA.

AC 088109;

DT 01-NOV-1998 (TREMUREL. 08, Created)

DT 01-NOV-1998 (TREMUREL. 08, Last sequence update)

DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)

DE GSD Protein.

GN GSD OR GTFD.

OS Mycobacterium avium subsp. silvaticum,

OS Mycobacterium paratuberculosis, and

OS Mycobacterium avium.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=44282, 1770, 1764;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=M.avium subsp. silvaticum, and M.paratuberculosis;

RA Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,

RT Hermon-Taylor J.;

RT "A low G+C content element in Mycobacterium avium subsp.

RT paratuberculosis and M. avium subsp. silvaticum with homologous genes

RT in M. tuberculosis.";

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

RU [2]

SEQUENCE FROM N.A.

RC SPECIES=M.avium; STRAIN=2151;

RA Eckstein T.M., Lambert M.L., Brennan P.J., Bellisle J.T., Inamine J.M.;

RT "Identification of a gene cluster involved in glycopeptidolipid

RT biosynthesis and of a gene cluster encoding daunorubicin resistance in

RT two strains of Mycobacterium avium serovar 2.";

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RU [3]

SEQUENCE FROM N.A.

RC SPECIES=M.avium; STRAIN=724;

RA Eckstein T.M., Brennan P.J., Inamine J.M., Bellisle J.T.;

RT "Identification of gene cluster involved in glycopeptidolipid

RT biosynthesis and of a gene cluster encoding daunorubicin resistance in

RT two strains of Mycobacterium avium serovar 2.";

RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ223832; CA11574.1; -
 DR EMBL: AJ223833; CA11578.1; -
 DR EMBL: AF143772; AAD4422.1; -
 DR EMBL: AF152999; AAD20376.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Transference.
 SQ SEQUENCE 266 AA; 30195 MW; F88A2754683F5A8B CRC64;

Query Match 100.0%; Score 1401; DB 2; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1,2e-114;
 Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTAPESTIIPFNAVTLQACLSIVGTYREVEVVLVDGSTDRTLDIANSFRELGS 60
 |||||
 DB 1 MTAPESTIIPFNAVTLQACLSIVGTYREVEVVLVDGSTDRTLDIANSFRELGS 60
 |||||
 QY 61 RLTVHSGPDGPDGYDAMNRGVATGEWVLFAGADDTLYEPTTLAQVAALFGDHAASHLVY 120
 |||||
 DB 61 RLTVHSGPDGPDGYDAMNRGVATGEWVLFAGADDTLYEPTTLAQVAALFGDHAASHLVY 120
 |||||
 QY 121 GDVNRSTKSRHAGPFDRLLEFETNLCHQSIFFYRRELFDGIGPYNLRYRMADMFENIR 180
 |||||
 DB 121 GDVNRSTKSRHAGPFDRLLEFETNLCHQSIFFYRRELFDGIGPYNLRYRMADMFENIR 180
 |||||
 QY 181 CPSNPLATRYMDVYISEYNDMTGFSMROGTKEFERKRLPMYFWAGWETCRMLAFKLD 240
 |||||
 DB 181 CPSNPLATRYMDVYISEYNDMTGFSMROGTKEFERKRLPMYFWAGWETCRMLAFKLD 240
 |||||
 QY 241 KENRRLATRLIRKAVSKERSAEP 266
 |||||
 DB 241 KENRRLATRLIRKAVSKERSAEP 266
 |||||

RESULT 2

QY 093025 PRELIMINARY; PRT; 297 AA.
 ID 093025

AC 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Putative glycosyl transferase Wbdl.
 GN Wbdl.
 OS Salmonella enterica.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=28901;
 RN [1]
 RP SEQUENCE FROM N.A.
 Wang L., Reeves P.R.;
 "The E. coli O111 and S. enterica O35 gene clusters: Gene clusters
 encoding the same colitose containing O antigen are highly
 conserved.";
 RT Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AF285969; AAK83018.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Transference.
 SQ SEQUENCE 297 AA; 34136 MW; AAGC9B82D322AB4B CRC64;

Query Match 21.3%; Score 298.5; DB 2; Length 297;
 Best Local Similarity 30.6%; Pred. No. 4,2e-18;
 Matches 72; Conservative 47; Mismatches 81; Indels 35; Gaps 7;

QY 4 PVFSTIIPFNAVTLQACLSIVGTYREVEVVLVDGSTDRTLDIANSFRELGSRLV 63
 |||||
 DB 18 PLISITIAIFNSELVIANLQSVISQYKNIETIIMDGSSYDRTLDIANSFKDE---RIK 74
 |||||
 QY 64 VHSQPDGPDGYDAMNRGVATGEWVLFAGADDTLYEPTTLAQVAALFGDHAASHLVYGDV 123
 |||||
 DB 75 IYSEKDRGIYDAMNRKAVDLSGDWIAFISDSDVYHTDAITSYVKGALISNGAPVYTG-- 132
 |||||

QY 124 VNRSTKSRHAGP-----FDLRLLEFETNLCH-----QSIFRRELFDGIGPYN 166
 |||||
 DB 133 -----RTAHGPKNEISGSGSEWYNLKGFRFNYYKNLPLPMSATISRDFFKD-GRFD 186
 |||||
 QY 167 LRYRVNADMFENIRCF-----SNPALITRYMDVYISEYNDM-TGFSMROGTDE 214
 |||||
 DB 187 IKRLIVADMDPLRCLFKRKRENSPYFIDITPYVRMGYGVSTDISQITRE 241
 |||||

RESULT 3

QY 095520 PRELIMINARY; PRT; 297 AA.
 ID 095520

AC 095520
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Putative glycosyl transferase.
 GN Wbdl.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 STRAIN-STOKE W;
 RC MEDLINE=96060831; Pubmed-7590310;
 RA Bastin D.A., Reeves P.R.;
 RT "Sequence and analysis of the O antigen gene (rfb) cluster of
 Escherichia coli O111.";
 RL Gene 164:17-23(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 STRAIN-STOKE W;
 RC MEDLINE=98449835; Pubmed-9774562;
 RA Wang L., Curd H., Qu W., Reeves P.R.;
 RT "Sequencing of Escherichia coli O111 O-antigen gene cluster and
 identification of O111-specific genes.";
 RL J. Clin. Microbiol. 36:3182-3187(1998).
 DR EMBL: AF078736; AAD46731.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Transference.
 SQ SEQUENCE 297 AA; 33740 MW; AF1B86797C2B37CB CRC64;

Query Match 21.3%; Score 298; DB 2; Length 297;
 Best Local Similarity 27.7%; Pred. No. 4,7e-18;
 Matches 78; Conservative 52; Mismatches 86; Indels 66; Gaps 8;

QY 1 MTAPESTIIPFNAVTLQACLSIVGTYREVEVVLVDGSTDRTLDIANSFRELGS 60
 |||||
 DB 15 LDAPLVIIITATVSELDIACLOSVTNOSYKNIETIIMDGSSDRTLDIANSFRD--- 71
 |||||
 QY 61 RLTVHSGPDGPDGYDAMNRGVATGEWVLFAGADDTLYEPTTLAQVAALFGDHAASHLVY 120
 |||||
 DB 72 RIKVISEKDRGIYDAMNRKADSLIGDWVAFIGSDVYHTDAITSIMKGVANSGAPVY 131
 |||||
 QY 121 GDVNRSTKSRHAGP-----FDLRLLEFETNLCHQSI-----FYRREL 158
 |||||
 DB 132 G-----RTAHGPDNNISGSGSEWYNLKGFRFNYYKNLPLPMSATISRDFFRNER 184
 |||||
 QY 159 FDGIGPYNLKRYVADMFENIRCF-----SNPALITRYMDVYISEYNDM-TGFSMROGT 211
 |||||
 DB 185 FD-----IKRLIVADMDPLRCLFKRKRENSPYFIDITPYVRMGYGVSTDISQVKT 238
 |||||
 QY 212 DKFERKRLPMYFWAGWETCRMLAFKDKENRRLATRLI 253
 |||||
 DB 239 TLE-----SFYRKKNNSCLNIQIL 259
 |||||

RESULT 4
 QY 08VTXO PRELIMINARY; PRT; 265 AA.
 ID 08VTXO
 AC 08VTXO

QY 173 ADMDFNIRCSNPALITRYMDVYISEYNDMTGFSM----RQGTDKFERKRLPMYF 223
 DB 174 ADEFFHFRWFKNKIKIKHSAVFVTLV-DPENGVSAPQPKNRKRIAEYOKIKKKYF 227

RESULT 7

OBVTW5 PRELIMINARY: PRT: 265 AA.

AC 08VWTW5; 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 31.4 kDa protein.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxID=173;
 RN NCBI_TaxID=173;
 RP SEQUENCE FROM N.A.
 RC STRAIN=SMITH.
 RX MEDLINE=21369626; PubMed=11476985;
 RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;
 "Genetic differences among the LPS biosynthetic loci of serovars of
 Leptospira interrogans and Leptospira borgpetersenii.";
 FEMS Immunol. Med. Microbiol. 31:73-81(2001).
 DR EMBL: AF316564; AAL49454.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 265 AA; 31395 MW; F078BE6F7A3A559D CRC64;

Query Match 19.6%; Score 275; DB 2; Length 265;
 Best Local Similarity 32.3%; Pred. No. 4.2e-16;
 Matches 76; Conservative 40; Mismatches 91; Indels 28; Gaps 6;

OY 4 PVFSIIPTFNAAVTLQACLSIGVQTYREVEVVLVDGSTD-----RTLDIANSFRP 56
 DB 6 PFIISITITNLANLEGLKRTLESVKQSYTNFELIVDGGSTGSEFYLSKNSLDIKKFIS 65
 OY 57 ELGSRIVVHSGPDDGPDYDAMNRGVATGEWVLFAGADDTLYEPTTLAQVAALFGDHAAS 116
 DB 66 E-----KDKGIYNAQNGKISLKEGYLVFLNAGDTLLQKNLLLEISKFLDDV-- 113
 OY 117 HUYGDVYMRSTKSRAHGFDDLLF---ETNLCHQSIFRRELFDGIGPYNLRYRW 172
 DB 114 DLVYGNILDSKDNGLIERKYPDLNLFYWSIKSLCHQAVFIKSLFDLYGYNEEYLF 173
 OY 173 ADMDFNIRCSNPALITRYMDVYISEYNDMTGFSM----RQGTDKFERKRLPMYF 223
 DB 174 ADEFFHFRWFKNKIKIKHSAVFVTLV-DPENGVSAPQPKNRKRIAEYOKIKKKYF 227

SULT 8

OBVTW9 PRELIMINARY: PRT: 265 AA.

AC 08VWTW9; 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 31.4 kDa protein.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxID=173;
 RN NCBI_TaxID=173;
 RP SEQUENCE FROM N.A.
 RC STRAIN=LA1;
 RX MEDLINE=21369626; PubMed=11476985;
 RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;
 "Genetic differences among the LPS biosynthetic loci of serovars of
 Leptospira interrogans and Leptospira borgpetersenii.";
 FEMS Immunol. Med. Microbiol. 31:73-81(2001).
 DR EMBL: AF316560; AAL49454.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.

KW Hypothetical protein.
 SQ SEQUENCE 265 AA; 31410 MW; DD9E916D09B77A9D CRC64;

Query Match 19.6%; Score 274; DB 2; Length 265;
 Best Local Similarity 32.9%; Pred. No. 5.1e-16;
 Matches 79; Conservative 40; Mismatches 91; Indels 30; Gaps 7;

OY 1 MTA--PVFSIIPTFNAAVTLQACLSIGVQTYREVEVVLVDGSTD-----RTLDIA 51
 DB 1 MFAKEPISITITNLANLEGLKRTLESVKQSYTNFELIVDGGSTGSEFYLSKNSLDI 60
 OY 52 NSRPPELGSRIVHSGPDDGPDYDAMNRGVATGEWVLFAGADDTLYEPTTLAQVAALFG 111
 DB 61 KRFISE-----KDKGIYNAQNGKISLKEGYLVFLNAGDTLLQKNLLLEISKFLD 110
 OY 112 DHASHLVYGDVYMRSTKSRAHGFDDLLF---ETNLCHQSIFRRELFDGIGPYNL 167
 DB 111 ODV--DLVYGNILDSKDNGLIERKYPDLNLFYWSIKSLCHQAVFIKSLFDLYGYNE 168
 OY 168 RYRWADMDFNIRCSNPALITRYMDVYISEYNDMTGFSM----RQGTDKFERKRLPMYF 223
 DB 169 EYLFAADFFHFRWFKNKIKIKHSAVFVTLV-DPENGVSAPQPKNRKRIAEYOKIKKKYF 227

RESULT 9

OBVT41 PRELIMINARY: PRT: 265 AA.

AC 08VLT41; 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 31.4 kDa protein.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxID=173;
 RN NCBI_TaxID=173;
 RP SEQUENCE FROM N.A.
 RC STRAIN=SALINEM, BALILCO, SMART, AND NAAM;
 RX MEDLINE=21369626; PubMed=11476985;
 RA de la Pena-Moctezuma A., Bulach D.M., Adler B.;
 "Genetic differences among the LPS biosynthetic loci of serovars of
 Leptospira interrogans and Leptospira borgpetersenii.";
 FEMS Immunol. Med. Microbiol. 31:73-81(2001).
 DR EMBL: AF316563; AAL49460.1; -
 DR EMBL: AF316556; AAL49446.1; -
 DR EMBL: AF316558; AAL49450.1; -
 DR EMBL: AF316562; AAL49458.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 265 AA; 31422 MW; E5539162FA877A9D CRC64;

Query Match 19.3%; Score 271; DB 2; Length 265;
 Best Local Similarity 32.3%; Pred. No. 9.3e-16;
 Matches 76; Conservative 40; Mismatches 91; Indels 28; Gaps 6;

OY 4 PVFSIIPTFNAAVTLQACLSIGVQTYREVEVVLVDGSTD-----RTLDIANSFRP 56
 DB 6 PFIISITITNLANLEGLKRTLESVKQSYTNFELIVDGGSTGSEFYLSKNSLDIKKFIS 65
 OY 57 ELGSRIVVHSGPDDGPDYDAMNRGVATGEWVLFAGADDTLYEPTTLAQVAALFGDHAAS 116
 DB 66 E-----KDKGIYNAQNGKISLKEGYLVFLNAGDTLLQKNLLLEISKFLDDV-- 113
 OY 117 HUYGDVYMRSTKSRAHGFDDLLF---ETNLCHQSIFRRELFDGIGPYNLRYRW 172
 DB 114 DLVYGNILDSKDNGLIERKYPDLNLFYWSIKSLCHQAVFIKSLFDLYGYNEEYLF 173
 OY 173 ADMDFNIRCSNPALITRYMDVYISEYNDMTGFSM----RQGTDKFERKRLPMYF 223
 DB 174 ADEFFHFRWFKNKIKIKHSAVFVTLV-DPENGVSAPQPKNRKRIAEYOKIKKKYF 227

us-09-705-911-24.rsp

Best Local Similarity 31.8%; Pred. NO. 1.9e-15;
Matches 83; Conservative 42; Mismatches 93; Indels 43; Gaps 10;

```
OY      4 PVFSIIIPFNAAVTLQACISIVGOTREVVVLVDGGSTD-----RTLDIANSFP 56  
        | | | | | : | : | : | : | : | : | : | : | : | : | : |  
Db      6 KRISITITNLNLEBGLRITESKQSYTNFELIYDVGSTGSFEYLKSINLDIRKFTS 65  
  
OY      57 ELGSRLLVHSGPDDPYDAMNRGVAVTGEWVLFAGADTLYEPITLAOVAALFDHAAS 116  
        | | | | | : | : | : | : | : | : | : | : | : | : | : |  
Db      66 E-----KDKRIYANQNGIKLSKGELYLVFLNAGALLQKVLSISIKFLQDI-- 113  
  
OY      117 HLVTGDVMVRSTKSRHACPEDL--DRLE---ETNLCHOSIFYRRLEPDGIGYNNRY 169  
        | | | | | : | : | : | : | : | : | : | : | : | : | : |  
        |||||:::~DSKDGHGIIEKKYPEDRLNYFWMSIKSLCHQATFFIRKNLFEDLYGYNNERY 170  
        | | | | | : | : | : | : | : | : | : | : | : | : | : |  
OY      170 RYMAMDENINRCSPALITRYMDVVISXYNDMTGSM----ROGDKDEFRRLPMYFYV 225  
        | | | | | : | : | : | : | : | : | : | : | : | : | : |  
Db      171 LPAALFEPPHRHWENKNKIYKAHPVILY-DPNGVSAOPKKAKRKRIAEYKRIRKKYPT 229  
        | | | | | : | : | : | : | : | : | : | : | : | : | : |  
OY      226 AGWETCRMLAFLDKDENRRL 246  
        | | | | | : | : | : | : | : | : | : | : | : | : | : |  
Db      230 -W-----AYVNKLNSYL 241
```

RESULT 12

ID	SEQUENCE	248 AA	PRT:	248 AA.
AD	09EXY4:			
AC	09EXY4:			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)			
DE	Putative glycosyl transferase.			
DE	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=M1182;			
RC	MEDLINE=21135136; PubMed=11238967;			
RA	Jensen S.O., Reeves P.R.;			
RT	"Molecular evolution of the GDP-mannose pathway genes (manB and manC)			
RT	in <i>Salmonella enterica</i> ."			
RL	Microbiology 147:599-610 (2001).			
DR	EMBL; AY012189; AAG41752.1;			
DR	InterPro; IPR001173; GLYcos transf_2.			
DR	Pfam; PF00535; GLYcos transf_2.1.			
KW	transferase			
SO	SEQUENCE	248 AA; 28560 MW; C01953610C8DEPAD CRC64;		

	Best Local Similarity	34.0%; Pred. No. 2.1e-15;	
Matches	67; Conservative	37; Mismatches	80; Indels
		13; Gaps	
OY	7	SIITTFNAAVYLQACISGIVGTYYREVEVVLYDGGSTDRTDIANSFRELGSRLV-VH	65
		: : - : : : : :	
Db	4	SIIITTYNSEKLRMTLESIELQYSDIERYIIIDGSTNTLKIIN---EVSTRTRCL	59
OY	66	SGPDGGPYAMNRGVATANGEMWFLTGADDTLEYPTTLQAVALFGDHASHLVGDVYM	125
		: : : : : : : : : : :	
Dd	60	SSDKGIYDALNKCIWLSTGDIIGFHSDDVLARPIIEITYGRFHETRAD--VVGDLVF	118
OY	126	RS-----TKSRHAGPFDDLRLLEFTNLCHOSIFYRRELFDGIGNYLRVMADMEN	178
		: : : : : : : : : : : :	
Dd	119	FKNQIDIKIRYWRSGPFRSKLSLGMAPHPSPFYMRRELYKNDGYDFDSLRIADYDQM	178
OY	179	IRCFSPNALITRYMAY 195	
		: : : : : : : : : : : : :	
Dd	179	VVLKRDDIKVSYPQV 195	

ID Q9ZGK3 PRELIMINARY; PRT; 251 AA.
 AC Q9ZGK3:
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Hypothetical 29.0 kDa protein.
 OS Leptospira borgpetersenii.
 OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxId=174;
 RX MEDLINE=93936580; PubMed=10458921;
 RA Kalamahetti T., Bulach D.M., Rajakumar K., Adler B.;
 RT Genetic organization of the lipopolysaccharide O-antigen biosynthetic
 RT locus of *Leptospira borgpetersenii* serovar Hardjovovis.";
 RL Microb. Pathog. 27:105-117(1999).
 DR EMBL: AF078135; AAD12966.1;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 251 AA; 29019 MW; 4C94B00C35D00EDC CRC64;

Query Match 18.2%; Score 255.5; DB 2; Length 251;
 Best Local Similarity 29.1%; Pred. No. 2e-14;
 Matches 67; Conservative 51; Mismatches 93; Indels 19; Gaps 8;
 QY 4 PVESIIPTFNAAVTLQACISIVGQ-TYREVEVVLVDGSGTDRFLDIANSFRELGSRL 62
 DB 6 PKISIVTINVDRLGKLEKTLISLRNICEPQIYIIIDAASKGCSLSVIOKTSDFSL 65
 QY 63 VVHSGPDDGPYDAMNKGCVATGEVWFLGADDTLYEPTTLAQVAFLGDHAAHLVYGD 122
 DB 66 ---SEPDLIYVGONKGIINSKGEYILFLNSGPTLASSENTLTETLSF---ELSSDLIYGD 119
 QY 123 VVMRSK-----SRHAGPFDRLLEFETNLCHOSIFRRELFDGIGPYNLRYVWADMDF 177
 DB 120 MLIESKGNLRIGROPVMTLSHLTLDT-IMHPACILKRKLFRYGLYDINFRIVADYEF 178
 QY 178 NIRCFSNPALITRYMDVVISSEYNDMTGFSMRQGTD---KEFRRLPMYF 223
 DB 179 WLKVPF-AGVSTKTYIPVFSQFN-LQGLSSAPQNSFLDQRRKKAQSLYF 226

RESULT 14
 Q9EXZ1 PRELIMINARY; PRT; 254 AA.
 ID Q9EXZ1:
 AC Q9EXZ1:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Putative glycosyl transferase.
 OS Salmonella enterica subsp. enterica.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 NCBI_TaxId=59201;
 RX MEDLINE=21135136; PubMed=11238967;
 RA Jensen S.O., Reeves P.R.;
 RT "Molecular evolution of the GDP-mannose pathway genes (manB and manC)
 RT in *Salmonella enterica*."
 RL Microbiology 147:599-610(2001).
 DR EMBL: AY012186; AAG41745.1;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Transferase.
 SQ SEQUENCE 254 AA; 29275 MW; 39759CD9576DB137 CRC64;

Query Match 18.2%; Score 255; DB 2; Length 254;
 Best Local Similarity 28.1%; Pred. No. 2.2e-14;
 Matches 74; Conservative 52; Mismatches 89; Indels 48; Gaps 8;

QY 7 SIIIPTFNAAVTLQACISIVGQ-TYREVEVVLVDGSGTDRFLDIANSFRELGSRLVHVS 66
 DB 4 SIIATATNSODTLKDTLTSVQDDYPDLEYIIIDGSSDQTLQILKKSTFVKCLII---S 60
 QY 67 GPDGPPDAMNKGCVATGEVWFLGADDTLYEPTTLAQVA-ALGDHAAHLVYGDVYM 125
 DB 61 EKDGITLDALNKGVKLAGDVGFTSHSDIILASQVLSINVKFRGNFA--DIYGDILF 118
 QY 126 RSTKSR-----HAGPFDRLLEFETNLCHOSIFRRELFDGIGPYNLRYVWADMDFN 178
 DB 119 VDRKRPDIIRHFWISGEEKSKLRYGNAPRPHAPYIRRELTKKGCPELRIKXIADYDM 178
 QY 179 IRCSFNPALITRYMDVVISSEYNDMTGFSMRQGTDEFRRLPMYFWAGMETCRMLAFL 238
 DB 179 LR-----LLAVPYQII---TYPEFVYMR-----LGGESTKLNNALL 213
 QY 239 KDKF-----NRRLAETR 251
 DB 214 STKEIIAMRRKHNIMQVAIIETR 236

RESULT 15
 Q9S4F9 PRELIMINARY; PRT; 263 AA.
 ID Q9S4F9:
 AC Q9S4F9:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Hypothetical 30.4 kDa protein.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
 NCBI_TaxId=173;
 RX MEDLINE=99403407; PubMed=10474199;
 RA De la Pena-Moctezuma A., Bulach D.M., Kalamahetti T., Adler B.;
 RT Comparative analysis of the LPS biosynthetic loci of the genetic
 RT subtypes of serovar Hardjo: *Leptospira interrogans* subtype
 RT Hardjoprajitno and *Leptospira borgpetersenii* subtype Hardjovovis.";
 RL FEMS Microbiol. Lett. 177:319-326(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA De la Pena-Moctezuma A., Bulach D.M., Kalamahetti T., Adler B.;
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF144879; AAD52183.1;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 263 AA; 30417 MW; EAD8DD59810E806 CRC64;

Query Match 18.2%; Score 254.5; DB 2; Length 263;
 Best Local Similarity 30.0%; Pred. No. 2.6e-14;
 Matches 69; Conservative 48; Mismatches 94; Indels 19; Gaps 8;
 QY 4 PVESIIPTFNAAVTLQACISIVGQ-TYREVEVVLVDGSGTDRFLDIANSFRELGSRL 62
 DB 6 PKISIVTINVDRLGKLEKTLISVRSQICLDQIEYIIIDAASKGDSLSIOGESHFSFWL 65
 QY 63 VVHSGPDDGPYDAMNKGCVATGEVWFLGADDTLYEPTTLAQVAFLGDHAAHLVYGD 122
 DB 66 ---SEPDLIYVGONKGIINSKGEYILFLNSGPTLASSENTLTETLSF---ELSSDLIYGD 119
 QY 123 VVMRSK-----SRHAGPFDRLLEFETNLCHOSIFRRELFDGIGPYNLRYVWADMDF 177
 DB 120 MLIESKGNLRIGROPVMTLSHLTLDT-IMHPACILKRKLFRYGLYDINFRIVADYEF 178
 QY 178 NIRCFSNPALITRYMDVVISSEYNDMTGFSMRQGTD---KEFRRLPMYF 223
 DB 179 WLKVPF-AGVSTKTYIPVFSQFN-LQGLSSAPQNSFLDQRRKKAQSLYF 226

Search completed: April 17, 2003, 17:27:55

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:21:00 ; Search time 11 seconds

(without alignments)
1002.973 Million cell updates/sec

Title: US-09-705-911-24

Perfect score: 1401
Sequence: 1 MTAFVFSIIPTFNAAVTLO.....ALRRLIRKAVKSKERSAEP 266

Scoring table:

BIOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	788	56.2	256	1 Y757_MYCTU	O50459 mycobacteri
2	202.5	14.5	262	1 Y714_MYCTU	P71793 mycobacteri
3	190.5	13.6	322	1 YS86_AMASP	P22639 anabaena sp
4	172.5	12.3	344	1 Y1BD_ECOLI	P11290 escherichia
5	168	12.0	250	1 Y868_HAEIN	O57022 haemophilus
6	165	11.8	323	1 YF78_HAEIN	O57287 haemophilus
7	147.5	10.5	248	1 WCAE_ECOLI	P71239 escherichia
8	146.5	10.5	1275	1 REBC_MYXXA	O50864 myxococcus
9	141	10.1	301	1 AMSB_ERWAM	O46632 erwinia amy
10	139	9.9	294	1 YG96_HAEIN	O48214 haemophilus
11	138.5	9.9	268	1 YMOF_BACSU	P39614 bacillus su
12	136	9.7	441	1 YCDO_ECOLI	P75905 escherichia
13	134	9.6	342	1 EXOU_RHIME	P33700 rhizobium m
14	134	9.6	348	1 EXOU_RHIME	P33697 rhizobium m
15	131.5	9.4	258	1 MAOE_RLEEN	O9X690 klebsiella
16	129.5	9.2	299	1 YAGI_RHISN	P55465 rhizobium s
17	127.5	9.1	299	1 Y025_MYCPN	P75086 mycoplasma
18	127.5	9.1	900	1 GGAB_BACSU	P46918 bacillus su
19	124.5	8.9	256	1 SP5A_BACSU	P39621 bacillus su
20	124.5	8.9	319	1 YF18_MYCTU	O50590 mycobacteri
21	124.5	8.9	446	1 GGAA_BACSU	P46917 bacillus su
22	124	8.9	298	1 Y025_MYCPN	P47271 mycoplasma
23	122.5	8.7	299	1 Y060_MYCPN	P75042 mycoplasma
24	119.5	8.5	426	1 CGED_BACSU	P42092 bacillus su
25	118.5	8.5	297	1 Y060_MYCPN	P47306 mycoplasma
26	113	8.1	403	1 YCOA_SYNP7	P43460 synechococc
27	109.5	7.8	333	1 RBBV_SALTY	P24401 salmoneilla
28	108	7.7	290	1 YAS7_MERTH	O38407 methanococc
29	105.5	7.5	241	1 DPM1_DROME	O9Y1U7 desoriphila
30	104	7.4	279	1 WCAE_ECOLI	P77414 escherichia
31	102.5	7.3	267	1 YG95_HAEIN	O48215 haemophilus
32	102	7.3	346	1 YF20_MYCTU	O50587 mycobacteri
33	101	7.2	340	1 REB4_ECOLI	P13961 escherichia

34	100	7.1	343	1 REP7_ECOLI	O52221 escherichia
35	100	7.1	343	1 REP2_ECOLI	P18023 escherichia
36	99.5	7.1	314	1 REBN_SALTY	P26403 salmoneilla
37	99.5	7.1	426	1 NODC_RHIME	P03431 rhizobium m
38	94.5	6.7	398	1 YAIR_ECOLI	O47536 escherichia
39	94	6.7	395	1 NODC_AZOCA	O07755 azorhizobiu
40	94	6.7	419	1 HASA_STRPY	O34865 streptococc
41	93.5	6.7	236	1 DPM1_SCHPO	O14466 schizosacch
42	92.5	6.6	260	1 DPM1_HUMAN	O60762 homo sapien
43	92	6.6	266	1 AMSB_ERWAM	O46635 erwinia amy
44	92	6.6	424	1 NODC_RHITO	P17862 rhizobium l
45	90.5	6.5	254	1 Y653_HAEIN	P44023 haemophilus

ALIGNMENTS

RESULT 1
ID Y757_MYCTU STANDARD: PRT; 256 AA.
AC O50459;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase RV2957 (EC 2.-.-.-).
GN RV2957 OR MT3031 OR MTCY349.31C OR U0002KC.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteriales (Class); Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN (1)
RP SEQUENCE FROM N.A.
RA Smith D. R., Robinson K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT *Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.*
RL Nature 393:537-544(1998).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Hatt D., Hickey E.,
RA Kolony A.J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterlind T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT *Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN (4)
RP -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
RC This SWISS-PROT entry is copyright. It is produced through a collaboration
RC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RC use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RC or send an email to license@sib-sib.ch).
DR EMBL; U00024; AAA50938.1; AUT_INIT.
DR EMBL; 283018; CAB05419.1; AUT_INIT.
DR EMBL; AE007125; AAK47357.1; -

```

DR TIGR: MT3031;
DR TubercList; Rv2957;
DR InterPro: IPR001173; Glycosyltransf_2.
DR Pfam: PF00535; Glycosyltransf_2.1.
DR Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 256 AA; 29012 MW; 90C6EC628C59CA57 CRC64;

Query Match
Best Local Similarity 56.2%; Score 788; DB 1; Length 256;
Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

OY 1 MTAPESTIIIFENAAVTIACLSIVGQTYREVEVYLVDSGSTDTLDIANSFRELGS 60
DB 1 MAAPESTIIIFENAAVTIACLSIVGQTYREVEVYLVDSGSTDTLDIANSFRELGS 60
OY 61 RLIVHSGPDDGPDYDAMNRGVATGEMVFLGADDTLYEPTTLAQAFAFLGDHAAHLVY 120
DB 61 RLIIHSDTQGYTDAMNRGVATGEMVFLGADDTLYEPTTLAQAFAFLGDHAAHLVY 120
OY 121 GDVNRSTKSRHAGPDDLRLFEETNLCHQSIFRRELFDGIGPYNLRYRWADMPENIR 180
DB 121 GDVNRSTKSRHAGPDDLRLFEETNLCHQSIFRRELFDGIGPYNLRYRWADMPENIR 180
OY 121 GDVNRSTKSRHAGPDDLRLFEETNLCHQSIFRRELFDGIGPYNLRYRWADMPENIR 180
DB 121 GDVNRSTKSRHAGPDDLRLFEETNLCHQSIFRRELFDGIGPYNLRYRWADMPENIR 180
OY 181 CENSNALITRYMDVVISSEYNDMTGFSMROGTDKEPRKRLPMTFWAGMETCRRLAFLKD 240
DB 181 CENSNALITRYMDVVISSEYNDMTGFSMROGTDKEPRKRLPMTFWAGMETCRRLAFLKD 240
OY 241 KENRRLATRLIRYK 256
DB 241 KENRRLATRLIRYK 256
OY 221 --STRGLRLVTLVLR 234
DB 221 --STRGLRLVTLVLR 234

RESULT 2
YF14_MYCTU STANDARD; PRT; 262 AA.
ID YF14_MYCTU
AC P71793;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase Rv1514c (EC 2.-.-.-).
GN Rv1514c OR MT1564 OR MTCY277.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE=96295987; PubMed=9634230;
RC Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RC Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekoa F.,
RC Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RC Davies R., Devlin K., Pettell T., Gentles S., Hamlin N., Holroyd S.,
RC Hornby T., Jagers K., Krogh A., McLean J., Moule S., Murphy J.,
RC Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RC Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RC Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RC "Deciphering the biology of Mycobacterium tuberculosis from the
RC complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RC Fletschman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RC Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RC Kolonay J.F., Nelson W.C., Omayam L.A., Ermolaeva M.D., Salzberg S.L.,
RC Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RC Bishai W.;
RC "Whole genome comparison of Mycobacterium tuberculosis clinical and
RC laboratory strains."
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 279701; CAB02028.1;
CC EMBL: AEO07023; AKK45831.1;
CC TIGR: MT1564;
CC TubercList; Rv1514c;
CC InterPro: IPR001173; Glycosyltransf_2.
CC Pfam: PF00535; Glycosyltransf_2.1.
CC Hypothetical protein; Transferase; Glycosyltransferase;
CC Complete proteome.
SQ SEQUENCE 262 AA; 28965 MW; 6B29BF8D31923E75 CRC64;

Query Match
Best Local Similarity 14.5%; Score 202.5; DB 1; Length 262;
Matches 68; Conservative 36; Mismatches 90; Indels 17; Gaps 9;

OY 2 TAPESTIIIFENAAVTIACLSIVGQTYREVEVYLVDSGSTDTLDIANSFRELGS 60
DB 3 SAPESTIIIFENAAVTIACLSIVGQTYREVEVYLVDSGSTDTLDIANSFRELGS 61
OY 61 RLIVHSGPDDGPDYDAMNRGVATGEMVFLGADDTLYEPTTLAQAFAFLGDHAAHLV 119
DB 62 --YWOSEPDGGRYDAMNRGVATGEMVFLGADDTLYEPTTLAQAFAFLGDHAAHLV 118
OY 120 YG---DVNRSTKSRHAGPDDLRLFEETNLCHQSIFRRELFDGIGPYNLRYRWADW 175
DB 119 WGFMDRLVGLDRVGRIPESLKRFLAGKQVYVQAASFFGSILVAKIGDLDGIADQ 178
OY 176 DENIRCESNPALITRYMDV--VISSEYNDMTG 204
DB 179 EFTLRA---ALVCEPVTIRCVLCEP-DTIG 204

RESULT 3
YS86_ANASP STANDARD; PRT; 322 AA.
ID YS86_ANASP
AC P22639;
DT 01-AUG-1991 (Rel. 19, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase alr2836 (EC 2.-.-.-).
GN ALR2836.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21595285; PubMed=11759840;
RC Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RC Kishida Y., Iriuchi M., Ishikawa A., Kawashima K., Klmura T.,
RC Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RC Yasuda M., Tabata S.;
RC "Complete genomic sequence of the filamentous nitrogen-fixing
RC cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213 (2001).
RN [2]
RP SEQUENCE OF 1-131 FROM N.A.
RC MEDLINE=90264305; PubMed=2111805;
RC Holland A., Wolk C.P.;
RC "Identification and characterization of beta, a gene that acts early
RC in the process of morphological differentiation of heterocysts."
RL J. Bacteriol. 172:3131-3137 (1990).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```


DR InterPro: IPR001173; glycos_transf_2,
PFam: PF00535; glycos_transf_2; 1.
KW Hypothetical protein; Transferase; glycosyltransferase.
KW Complete proteome.
SQ SEQUENCE 250 AA; 28915 MW; A5DB220129782E98 CRC64;

Query Match	12.0%;	Score 168;	DB 1;	Length 250;
Best Local Similarity	25.8%;	Pred. NO. 3e-08;		
Matches 57;	Conservative 39;	Mismatches 83;	Indels 42;	Gaps 7;

Oy 1 MNPVFSIIPTFENNAVTLQACLSGIYQOTRYREVVLVDGGSTORTLDIANSFRPEIGS 60
| : |||| : || : : : || : |||| : |||| : || :
Dd 1 MNMPLISIMPVYNACYLNGCILSCINQSYQNEILILDGSGTDSKEIINNIIIDDKR 60

Qy 61 RLTVHSGPDDGPPYDAMNRGVATGEMVLEFLGADDTLYEPTTLAQVAAFLGDH-AAHLV 119
 : : | | | | : : : : | | : : : : |
 Db 61 VKLEFPTNGPAAARNIGLEKAQGDVTFLLSDDFIANDKLEKQNLMTLQNHLLVMTHTGN 120

QY 120 Y-----GDVVMRSKSR-----HAGPFDLRLFE-----TNLC-----149
| : : | : : |
Db 121 YAFCDEGNGIKLVTSKKIDYLTLLGGNQFKIMTVLEVESIKILRFNPKIHEDYAFFL 180C

Db

```
150 -----QSIYRREL--FDGIGFVNLRIKRW--ADMDENI 179
      ||| : ||| :: |||
181 DCLKEVQSIILYSHQASSFVRIGKAVSSSNKFKSLMTFTNI 221
```

RESULT 6
YF78_HAEIN

AC	Q57287; 005077;
DT	15-JUL-1998 (Rel. 36, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)

DE Putative glycosyl transferase H11578 (EC 2.-.-.-)
GN H11578.
OS Haemophilus influenzae.

```
OC Haemophilus.
OX NCBI_TaxID=727;
RN '[1]
```

RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; Pubmed=7542800;
RA Fleischmann R.D., Adams M.D., White

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 Venter J.C.;

RT Rd. #;
RL Science 269:496-512(1995).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CS

CC This SWISS-PROT entry is copyright. It is produced through
CC between the Swiss Institute of Bioinformatics and the EMBL
CC the European Bioinformatics Institute. There are no restric-

cc modified and this statement is not removed. Usage by and for commercial
cc entities requires a license agreement (See <http://www.isb-sib.ch/announce>
cc or send an email to license@isb-sib.ch).

DR EMBL; U32832; AAC23227.1; -.
DR TIGR; H11578; -.
DR InterPro; IPR001173; Glycos_transf_2
DR Pfam; PF00035; Glycos_transf_2; 1

KM	Hypothetical protein; Transferase; Glycosyltransferase
KW	Complete proteome.
SO	SEQUENCE 323 AA; 37680 MW; 7CBBC2681039AB5B4 CRC64;

Query Match	11.8%	Score	165;	DB	1;	Length	323;
Best Local Similarity	22.5%	Pred. No.	7.9e-08;				
Matches	48;	Conservative	51;	Mismatches	86;	Indels	28;
						Gaps	9;

[illegible]

QY 64 VHSGGDD-GPFDAMNRGVAVAGEWLFLGADPTLYEPITLLAQVAALF--GDHAASHLVY 120C

DB 64 IISNRYNLGFINSLNIGLGFCSFKYFARDADD-IAPSWIEKITYVLEKNDHITAGSY 1222

```

QY 121 GDVVRSTKSNHAGPDDLRL-----LEFNLCHSIFR-----RE---LF 159
      ::: : | :: | | : |
Db 123 LEIIVEKEGIIIGQYKTGDIWKNPDLHNDICEAMLFYNP1HNNTIMIRANYREHKLIF 182

```

```

QY      160  DGIGPYNLRRVWADWDENIRCFNSN-PALITRY 191
          :  ||  |::|  :  ||  :  :|
Db      183  NKDYPAEDYKFWSEVS-RLGCLANYPEALVKY 214

```

RESULT 7
WCAE_ECOLI

AC	P71239; P76384;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)

DE Putative colanic acid biosynthesis glycosyl transferase wcae
GN WCAE OR B2055.
OS *Escherichia coli*.

OC	Escherichia.
OX	NCBI_TaxID=562;
RN	[1]

RC STRAIN=K12;
RX MEDLINE=96326333; PubMed=8759852;
RA Stevenson G., Andrianopoulos K., H
P

RT for production of the extracellular polysaccharide colanic acid.";
RL J. Bacteriol. 178:4885-4893(1996).
RN [2]

RC STRAIN-K12;
RA Reeves P.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; Pubmed=9278503;

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

RL Science 277:1453-1474(1997).
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF THE SLIME POLYSACCHARIDE
CC COLANIC ACID.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use.

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U38473; AAC77840.1; -.
DR EMBL; AE000295; AAC75116.1; -.
DR EcoGene; EG13573; wcae.
DB JGI; JGI-001172; clust4

DR LUCERPIO; 1PKU11/3; GLYCOS_TRANSI_2


```

DB 123 YGQVNIIDGSLKVMPLKPVGEHSIAEYLFQCYG-FIOTSTIVLKREDAEIRED--- 180
QY 143 FETNICHOSIETFRRELFDGIPYLNRYWADMDFNIRCFSPNALTIRYMDVY--ISEVN 200
DB 181 -ERYIRHQ-----DYDL---CIRADKLGFKFMVINOPIALANYH 213
QY 201 DMTGF-SMGOCTDKER-----RKLPMTFVYVAGETCRRLMA-- 236
DB 214 MYTRGSOHKGSGSVKYSLEFLMDAMKPHLTRRDVYTYKAYKPLPYRKMDG-KSLQASLSFA 272
QY 237 ---FLDKENR-----RL--ALRTLL 252
DB 273 RYFFLELNKDRNDELKRLMKLRTLL 298

RESULT 10
YQ96_HAEIN STANDARD; PRT; 294 AA.
AC 048214; 005082;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
Putative glycosyl transferase H11696 (EC 2.---).
H11696
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
NX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=42;
RA McLaughlin R., Abu Kwaik Y., Young R., Spinoia S., Apicella M.;
RT "Characterization and sequence of the lsg locus from Haemophilus
RT influenzae.";
RL Submitted (JUN-1992) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleisichmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitchugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fline L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RA Rd.";
RT Science 269:496-512(1995).
-i SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M94855; AAA24982.1; -
CC EMBL; U32842; AAC23342.1; -
CC TIGR; H11696; -
DR InterPro; IPR001173; Glycos transf_2.
DR Pfam; PF00535; Glycos_transf_2.1.
KM Hypothetical protein; Transferase; Glycosyltransferase;
KM Complete proteome.
FT CONFLICT 38 C -> Y (IN REF. 1).
FT CONFLICT 48 S -> R (IN REF. 1).
FT CONFLICT 70 V -> I (IN REF. 1).
FT CONFLICT 74 T -> A (IN REF. 1).
FT CONFLICT 88 C -> R (IN REF. 1).

```

```

FT CONFLICT 97 97 V -> I (IN REF. 1).
FT CONFLICT 106 106 N -> D (IN REF. 1).
FT CONFLICT 152 152 I -> V (IN REF. 1).
FT CONFLICT 185 185 P -> S (IN REF. 1).
SQ SEQUENCE 294 AA; 33646 MW; 8330F081BFFA18B CRC64;

Query Match 9.98; Score 139; DB 1; Length 294;
Best local similarity 23.0%; Pred. No. 1,7e-05;
Matches 44; Conservative 40; Mismatches 93; Indels 14; Gaps 5;

QY 5 VFSTIITFENAAVTLQACISIVGQITREVVAVLVDSGTIRTDIANSPFELGSLVY 64
DB 1 MLSTIVSYNKRKAEVPALESITLTOOTSNSFVYIIVDDCSKERYV-VEQSYSEFPV---TVI 56
QY 65 HSGPDGPPYDAMNRCGVATGEVWFLGADDTLVEPTTLAQAVALFGHAAASHLVY---- 120
DB 57 RNEFNQGAESRNVGARTSKGDMFLFLD-DDDCRPEKCEVLOYIEQNPINITYHPAK 115
QY 121 ----GDVVMSTKSRHAGPFDDLRLFEETNLCHO-SIFVRELPDGIQPNLRYRWADW 175
DB 116 CEMVNEGFYVYTOPIEPOEISTERILLANKIGCHPMIAIKKEMFLKIGLSTALRSLEDY 175
QY 176 DFNIRCFSPNRA 186
DB 176 DFLKLQLQEPS 186

RESULT 11
YQ96_BACSU STANDARD; PRT; 268 AA.
AC P39614;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase ywdf (EC 2.---).
GN YWDF OR IPA-56D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hall M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Prescan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Beszies P., Bilotin A., Borchert S.,
RA Brouillet R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani U.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
RA Entlin R.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Glaser P., Fujita Y., Fuma S., Gallizi E.J., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

```

RA Rieger M., Rivolta C., Rocca E., Rocche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretti A.,
 RA Viati A., Wambuit R., Wedler E., Wedler H., Weizenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.,
 RT The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.
 RL Nature 390:249-256(1997).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X73124; CAB51612.1; -
 DR EMBL: 269123; CAB51824.1; -
 DR PIR: S39711; S39711.
 DR Subtilisin; BG10602; ywdf.
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00053; Glycos_transf_2; 1.
 DR Hypothetical protein; Transferase; Glycosyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 268 AA: 30616 MW: DD6428F7016EC9B3 CRC64;
 Query Match 9.9%; Score 138.5; DB 1; Length 268;
 Best Local Similarity 22.5%; Pred. No. 1.7e-05;
 Matches 61; Conservative 52; Mismatches 115; Indels 43; Gaps 10;
 QY 7 STIIPENAAVTLQACLSIVGQTYREVEVYLV-DGSGTRTIDIANSPREPGLSRLVH 65
 DB 4 STIIVYNNIPALCELLLESISRTQMPYEITIIYNDAG---ESVVPKALVPEL-PIAVIN 59
 QY 66 SGPDDGPDYAMNRGVANGVATGEVWFLFGADDTLYPEPTTIAOVAFLGDHAASHLVYGVVA 125
 DB 60 LEKNSGVAARNAGVKEASGDCIM-LCDDDDFTTPGHEIMAK---EIEIADYVHSAEI 115
 QY 126 RSTKSHAGPFDRLLEFETNLCHQST-----FYRRELFGIGPYNLRYRWAD 174
 DB 116 VSEFEENGRYRYPVSRKLFATYADYEDMRVSTYVPSGMYRFLHDEIGYDADVHNYMD 175
 QY 175 WDFNICEFN-----PALITRYMDVYISEYNDMTGFSMROGTDEKFRK---RLPMY 222
 DB 176 WDFYLAADYRVKRVPCASVIYAFSDAGNOSADI-GAKRKQYLDRLSEKHGLGELP-- 232
 QY 223 FVWAGMETCRMLAFLKDEKRNRLALTRLI 253
 DB 233 -----TKNFAVLEPEPMKREKSENV 255
 RESULT 12
 ID YCDO_ECOLI STANDARD: PRT: 441 AA.
 AC P75905;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Hypothetical protein ycd0.
 GN YCDO OR B1022.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Berra N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.U.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-153(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: TO Y. PESTIS HMS LOCUS PROTEIN HMSR AND TO
 CC S. EPIDERMIDIS ICA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000204; AAC74107.1; -
 DR EMBL: D90739; BAA35803.1; -
 DR EcoGene: EG13863; ycd0.
 DR InterPro: IPR001173; Glyco. transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 5 25
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 331 351 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 395 415 POTENTIAL.
 SQ SEQUENCE 441 AA: 50765 MW: 04F5A53D72FEBAAB CRC64;
 Query Match 9.7%; Score 136; DB 1; Length 441;
 Best Local Similarity 19.9%; Pred. No. 5.4e-05;
 Matches 54; Conservative 41; Mismatches 101; Indels 76; Gaps 8;
 QY 4 PVFSIIPTFNAATVTLQACLSIVGQTYREVEVYLV-DGSGTRTIDIANSPREPGLSRLV 63
 DB 75 PVSIIITPCFNEKNEVETIIHAALAQRENEIVAVNDGSDKRALIDMAAQIPHLRY 134
 QY 64 VHSPPDGPDIYAMNRGVANGVATGEVWFLFGAD-----DTLYPEPTT----- 102
 DB 135 ITHAONOGKAIATLKTGAANAASEYLVICIDGALLDRDAAYIVPMALNPRVGAATGNPR 194
 QY 103 -----LAOVAFLGDHAASHLVYGDV-----VMSTKSRHAGPF----- 136
 DB 195 IRTSTLVGKIQVGEYSIIILIKRTQRIYGNVTVSGVIAFRSALAENVGWSDDMIT 254
 QY 137 -DIDRLLEFETNLCHQSIYR-----RELFGIGPYNLRY-----RYVA 173
 DB 255 EDID-ISMKQLQNMOTFEYERALCWILMPETLGLMKORLWMAOGAEVFLKNMTRILMR 313
 QY 174 WDFNICEFSPALITRYMDVYISEYNDMTGE 205
 DB 314 KENFRMW-----PLFEYCLTTIWAFTCLVGF 340
 RESULT 13
 ID EXOU_RHIME STANDARD: PRT: 342 AA.
 AC EXOU_RHIME

```

AC P33700;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Succinoglycan biosynthesis protein exou (EC 2.7.1.1).
GN EXOU OR RB1071 OR SMC20948.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=94042869; PubMed=8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT succinoglycan by Rhizobium meliloti."
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / S047;
RX MEDLINE=94162682; PubMed=8118055;
RA Becker A., Kleickmann A., Kuester H., Keller M., Arnold W.,
RA Puehler A.;
RT "Analysis of the Rhizobium meliloti genes exou, exov, exow, exot, and
RT exoi involved in exopolysaccharide biosynthesis and nodule invasion:
RT exou and exow probably encode glycosyltransferases."
RL Mol. Plant Microbe Interact. 6:735-744(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
CC SUCCINOGLYCAN (EPS I). NEEDED FOR THE ADDITION OF THE SIXTH SUGAR
CC (GLUCOSE). CATALYZES THE FORMATION OF A BETA-1,6 LINKAGE BETWEEN
CC THE FIFTH AND SIXTH SUGAR.
CC -1- PATHWAY: Exopolysaccharide biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: L20758; AAA16053.1;
DR EMBL: 222646; CAA80359.1;
DR EMBL: AL603645; CAC49471.1;
DR PIR: D49348; D49348.
DR InterPro: IPR001173; Glycosyltransf.2.
DR Pfam: PF00535; Glycosyltransf.2.1.
DR Transfaser: Glycosyltransferase; Exopolysaccharide synthesis;
KM Plasmid; Complete proteome.
SQ SEQUENCE 342 AA; 37017 MW; EAF55E0EBDA023BC CRC64;

```

```

QY 68 PDGPGYDAMRGVATGENTVLEFGADDTLYEPTTLAQVAALFg-DHAASHLYGVGYMR 126
DB 70 ENRGPAARHAAHAISHPLIGVLDADDFEF-PGRQLQLSQDQWDFADINIAIAQA 128
QY 127 STGRSHAGGP-DIDRL-----FETNCHOSI-----FYRRELFDSIG-PYNLR 169
DB 129 ATAHGRDRAPPPRLDLVGVFEGNISRGVRRGEIGFLKPLRRAPFLDQHGRLVETL 188
QY 170 RVNADMFNIRCFSPNLTIRYMDVVISSEYNDMTGFSMROCTDKEF-RKRLPMFVWAGW 228
DB 189 RLEDVDLYARALANGA---RYKIHSCGYAAVVRGNSLSGSHRTIDLRK-----Y 237
QY 229 ETCRRMLATLKDKENRRRLAR 249
DB 238 EADRAILAGSRLSSDAEAVR 258

RESULT 14
ID EXOU_RHIME STANDARD; PRT; 348 AA.
AC P33697;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Succinoglycan biosynthesis protein exou (EC 2.7.1.1).
GN EXOU OR RB1084 OR SMC20959.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=94042869; PubMed=8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT succinoglycan by Rhizobium meliloti."
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / S047;
RX MEDLINE=21396508; PubMed=11481431;
RA Becker A., Kleickmann A., Keller M., Arnold W., Puehler A.;
RT "Identification and analysis of the Rhizobium meliloti exoMOMP genes
RT involved in exopolysaccharide biosynthesis and mapping of promoters
RT located on the exoHKLAMP fragment."
RL Mol. Gen. Genet. 241:367-379(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / S047;
RX MEDLINE=94067019; PubMed=8246891;
RA Becker A., Kleickmann A., Keller M., Arnold W., Puehler A.;
RT "Identification and analysis of the Rhizobium meliloti exoMOMP genes
RT involved in exopolysaccharide biosynthesis and mapping of promoters
RT located on the exoHKLAMP fragment."
RL Mol. Gen. Genet. 241:367-379(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
CC SUCCINOGLYCAN (EPS I). NEEDED FOR THE ADDITION OF THE FIFTH SUGAR
CC (GLUCOSE). CATALYZES THE FORMATION OF A BETA-1,6 LINKAGE BETWEEN
CC THE FOURTH AND FIFTH SUGAR.
CC -1- PATHWAY: Exopolysaccharide biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 208.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:25:05 ; Search time 47 Seconds
(without alignments)
544.080 Million cell updates/sec

Title: US-09-705-911-24

Perfect score: 1401

Sequence: 1 MTAPEVFSIIPTFNAATLQ.....ALTRLIRYKAVSKERSAPP 266

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	788	56.2	275	2	B70670
2	238.5	17.0	247	2	AG0376
3	232.5	16.6	260	2	E71975
4	231.5	16.5	259	2	F64532
5	220	15.7	324	2	AB2190
6	207	14.8	321	2	AG2188
7	206.5	14.7	316	2	AE2189
8	205.5	14.7	298	2	B75096
9	204.5	14.6	330	2	AE2188
10	202.5	14.5	262	2	E70714
11	196.5	14.0	337	2	AG1920
12	196.5	14.0	367	2	G95948
13	192	13.7	336	2	AG168
14	190.5	13.6	322	2	AE2160
15	190.5	13.6	323	2	AD2189
16	190	13.6	248	2	C90984
17	190	13.6	248	2	F85829
18	189.5	13.5	333	2	AE2026
19	188.5	13.5	333	2	B97168
20	188	13.4	299	2	B83557
21	182	13.0	344	2	AC0974
22	178	12.7	356	2	S74766
23	177.5	12.7	343	2	AT2091
24	175	12.5	318	2	AG2189
25	174.5	12.5	333	2	G86651
26	174.5	12.5	623	2	ARI209
27	174.5	12.5	996	2	S76194
28	173.5	12.4	324	2	A69290
29	173.5	12.4	338	2	E91190

30	173.5	12.4	338	2	F86037	probable regulator
31	172.5	12.3	344	1	O3ECRH	hypothetical 40.5k
32	172	12.3	573	2	B89789	hypothetical prote
33	171	12.2	344	2	A70037	capsular polysacch
34	170.5	12.2	301	2	F95205	glycosyl transfera
35	170	12.1	354	2	H96021	probable glycosyl
36	169.5	12.1	774	2	AC1566	hypothetical prote
37	168.5	12.0	294	2	G71148	probable glycosyl
38	168	12.0	250	2	A64039	glycosyl transfera
39	168	12.0	318	2	AH2189	hypothetical prote
40	168	12.0	462	2	AH1566	galactosamine-cont
41	168	12.0	1013	2	AE1876	hypothetical prote
42	165.5	11.8	340	2	T44330	glycosyl transfera
43	165	11.8	303	2	AB4114	glycosyl transfera
44	165	11.8	323	1	H64130	hypothetical prote
45	164	11.7	334	1	G71153	

ALIGNMENTS

RESULT 1

B70670

hypothetical protein Rv2957 - Mycobacterium tuberculosis (strain H37RV)

N:Alternate names: u0002kc protein

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence.revision 17-Jul-1998 #text_change 20-Jun-2000

C/Accession: B70670; S73064

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen

A: Reference number: A70500; PMID:98295987; PMID:9634230

A: Accession: B70670

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-275 <COL>

A: Cross-references: GB:283018; GB:AL123456; NID:93261671; PIDN:CA805419.1; PID:91694

A: Experimental source: strain H37RV

R: Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, September 1994

A: Description: Mycobacterium tuberculosis cosmid tbcc2.

A: Reference number: S73053

A: Accession: S73064

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-275 <SMI>

A: Cross-references: EMBL:U00024; NID:9560506; PIDN:AAA50938.1; PID:9560517

C: Genetics:

A: Gene: Rv2957

C: Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 56.2%; Score 788; DB 2; Length 275;

Best Local Similarity 60.9%; Pred. No. 2.6e-64;

Matches 156; Conservative 24; Mismatches 54; Indels 22; Gaps 2;

QY	1	MTA	VF	FSII	PTFNAATLQACISIVGQYREVYVVDGSTMRTDIANSFPEIGS	60
DB	20	MAA	MF	SI	IIPTLVNAANVLPACLOSTIARQTCGDFELVVDGSDTETDIANIFAPNIGE	79
QY	61	RLV	VS	GPDD	GPYDAMNKGAVATGEVYLFAGADDTLEPTLQAOVAFLGDHAAHLYV	120
DB	80	RLI	HR	PD	GVYDAMNKGAVATGEVYLFAGADDTLEPTLQAOVAFLGDHAAHLYV	139
QY	121	GDV	VS	RS	TKSHAGPFDLRLFTETNLCQSIYFRRELFDGIGPYNLRYRWADMDENIR	180
DB	140	GDV	VS	RS	TKSHAGPFDLRLFTETNLCQSIYFRRELFDGIGPYNLRYRWADMDENIR	199
QY	181	CF	SN	PA	LITRIMDVVISYNDMTGFSMNGQTDKPKRLPMYFWAGWTCRMLAFLKD	240
DB	200	CF	SN	PA	LITRIMDVVISYNDMTGFSMNGQTDKPKRLPMYFWAGWTCRMLAFLKD	239

[illegible]

A: Molecule type: DNA

A: Residues: 1-323 <KUR>

A: Cross-references: GB:BA000019; PIDN:BA074766.1; PID:q17132161; GSPDB:GN00179

A: Experimental source: strain PCC 7120

C: Genetics:

A: Gene: alr3067

Query Match 13 6%; Score 190.5; DB 2; Length 323;

Best Local Similarity 22.2%; Pred. No. 1.1e-09; Mismatches 69; Indels 73; Gaps 6;

Matches 53; Conservative 44; Mismatches 69; Indels 73; Gaps 6;

QY 4 PVFSIIIPFNAAVTLQACISIVGQTYREVEVLVDGSGSTRTLDIANSFRPELGSRLV 63

DB 11 PLISVIIPYNGEKTIETIASVQHOTFLDIETIYINDGSTDNTFELVNRNQ---DNRLK 67

QY 64 VHSQPDGPGYDAMNRGVATGEMWLELGAD-----TL 97

DB 68 IFSYENGGLPVARNRGITFAVGQFIADADDLMTDKLELQFALQEYEPENGLAYSWTY 127

QY 98 YEPITLAQYAA-----FLGDHAASHLYGDVYMRSTKSRHAGPFDRLLEFETNLCHQSI 152

DB 128 YKFANEADSYADESNSFAGD-----VYAEELIKNFLQNGSNP----- 164

QY 153 FYRRELFQDGIQPPYNLRYRYVADMDENIRCS-----NPALIRRYMDVY 195

DB 165 LIRRAAIDSVGLFDPPLKSCEDMDFTLRLAKWQFALVKKAOIYROSPTAMTSKLDM 223

Search completed: April 17, 2003, 17:28:55
Job time : 56 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:20:40 ; Search time 75 Seconds

(without alignments)
472.596 Million cell updates/sec

Title: US-09-705-911-24

Perfect score: 1401
Sequence: 1 MFAVPSIIITPTFNAAVTLQ.....ALRRRLIRKAVSKERSAEP 266

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

arched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

```

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1401	100.0	266	18	AAW21774
2	1401	100.0	266	18	AAW21775
3	788	56.2	275	18	AAW21783
4	298	21.3	297	20	AAW8309
5	205.5	14.7	298	22	AAW6313
6	202.5	14.5	262	18	AAW21779
7	201.5	14.4	321	23	ABP36805
8	190	13.6	248	20	AAW8314
9	186.5	13.3	316	21	AAW54072
10	186.5	13.3	316	21	AAW43774

11	186	13.3	277	18	AAW21767
12	184.5	12.9	327	22	ABW47426
13	180.5	12.9	327	23	ABW26804
14	179	12.8	278	21	AAW68976
15	175	12.5	334	21	AAW68963
16	174.5	12.5	333	23	ABW53519
17	174.5	12.5	623	23	ABW48553
18	172	12.3	358	22	AAU34218
19	172	12.3	573	22	AAU34218
20	167.5	12.0	295	23	ABW27408
21	165.5	11.8	336	21	AAW54098
22	165.5	11.8	336	21	AAW54098
23	165	11.8	251	22	AAW90448
24	163	11.6	332	21	AAW68962
25	162.5	11.6	274	22	AAW90151
26	161.5	11.5	346	23	AAU72923
27	161	11.5	324	18	AAW14078
28	161	11.5	324	18	AAW22177
29	160.5	11.5	322	21	AAW54071
30	160.5	11.5	322	21	AAW43773
31	158.5	11.3	327	21	AAW54095
32	158.5	11.3	327	21	AAW43797
33	158	11.3	348	17	AAW91311
34	158	11.3	348	18	AAW06576
35	157.5	11.2	706	22	AAU33454
36	157.5	11.2	715	22	AAU34918
37	157	11.2	327	23	ABW48565
38	156.5	11.2	270	21	AAW97203
39	153	10.9	322	21	AAW68974
40	152.5	10.9	328	21	AAW81720
41	152	10.8	322	21	AAW68975
42	149.5	10.7	972	20	AAW06212
43	149.5	10.7	972	21	AAW43099
44	148.5	10.6	963	23	ABW54234
45	147.5	10.5	965	21	AAW96213

ALIGNMENTS

```

RESULT 1
AAW21774
11-MAR-1998 (first entry)
AC AAW21774;
XX
XX
DE Protein encoded by ORF F of GS region in M. avium subspecies silvaticum.
XX
XX
KW GS; pathogenesis island; pathogenic protein; mycobacterial disease;
KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
KW vaccine; inflammatory disease; sarcoidosis; Johne's disease.
XX
XX
OS Mycobacterium avium subspecies silvaticum.
XX
XX
PN WO9723624-A2.
XX
XX
PD 03-JUL-1997.
XX
XX
PF 23-DEC-1996; 96WO-GB03221.
XX
XX
PR 21-DEC-1995; 95GB-0026178.
XX
XX
PA (SGRO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
XX
XX
PI Doran T, Ford J, Hermon-Taylor J, Loughlin M, Miller D;
XX
XX
PI Sumar N, Tizard M;
XX
XX
PI WPI; 1997-351061/32.
XX
XX
PI N-PSDB; AAT74469.
XX
XX
PI New isolated pathogenicity island from mycobacteria - used to

```

Protein encoded by
EpsN, Lactococcus
Streptococcus poly
CpsIX protein whic
CpsIX protein whic
Lactococcus lactis
Lactococcus monocyt
Staphylococcus aur
Staphylococcus aur
Streptococcus poly
Enzyme EpsN, Lacto
Amino acid sequenc
C glutamicum prote
CpsIX protein whic
C glutamicum prote
Neisseria meningit
S. thermophilus exo
S. thermophilus exo
Enzyme Eps7 which
Amino acid sequenc
Enzyme Eps7 involv
Amino acid sequenc
N. gonorrhoeae gly
Neisseria polyglyc
Enterococcus faeca
Enterococcus faeca
Listeria monocytog
Campylobacter jej
CpsII protein whic
Streptococcus pneu
CpsIX protein whic
Pasteurella multoc
P. multocida hyalu
Lactococcus lactis
P. multocida chond

PT develop products for detection, diagnosis, prevention and treatment
of mycobacteria infections

PS Claim 1; Page 50; 62pp; English.

CC The present sequence represents a protein encoded by open reading frame
CC (ORF) F of a novel polynucleotide sequence designated "GS". GS is a
CC pathogenecity island of 8 kb of DNA comprising a core region of 5.75 kb
CC with multiple ORFs and an adjacent transmissible element of 2.5 kb. The
CC ORFs, and also the transmissible element, encode proteins which may be
CC linked to pathogenecity, such as providing receptors for cellular
CC recognition. GS was discovered and characterised using differential DNA
CC analysis technology. It is found within Mycobacterium paratuberculosis
CC and it has also been identified in Mycobacterium avium subspecies
CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
CC the intestine and Crohn's disease in humans. The protein products of the
CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
CC treating or preventing mycobacterial disease. In particular they can be
CC used as vaccines for inflammatory diseases such as Crohn's disease or
CC sarcoidosis in humans or John's disease in animals.

XX Sequence 266 AA;

Query Match 100.0%; Score 1401; DB 18; Length 266;
Best Local Similarity 100.0%; Pred. No. 7.3e-144;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MTAPEVSIITPFNAAVTLQACISIVGQTYREVEVVLVDGSGTDRITDIANSFRELGS 60

QY 61 RLTVHSGPDDGPRYDANNRGVATGEMVFLGADDTLYEPTTLAQAVALGDMASHLY 120

DB 61 RLTVHSGPDDGPRYDANNRGVATGEMVFLGADDTLYEPTTLAQAVALGDMASHLY 120

QY 121 GDVYMRSTKSRHAGFPDRLLEFETNLCHOSIFRYRELEFDGIGPYNLRYVADWDFNIR 180

DB 121 GDVYMRSTKSRHAGFPDRLLEFETNLCHOSIFRYRELEFDGIGPYNLRYVADWDFNIR 180

QY 181 CFSNPALITRYMDVYISEYNDMTGFSMRGTDKEFRRLPMYFWAGWETCRMLAFKLD 240

DB 181 CFSNPALITRYMDVYISEYNDMTGFSMRGTDKEFRRLPMYFWAGWETCRMLAFKLD 240

QY 241 KENRRALRTLRIRYKAVSKERSAEP 266

DB 241 KENRRALRTLRIRYKAVSKERSAEP 266

RESULT 2

AAW21775 standard; Protein; 266 AA.

AAW21775;

DT 11-MAR-1998 (first entry)

DE Protein encoded by ORF F of the GS region in M. paratuberculosis.

KW GS; pathogenecity island; pathogenic protein; mycobacterial disease;
cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
vaccine; inflammatory disease; sarcoidosis; John's disease.

OS Mycobacterium paratuberculosis.

PN WO9723624-A2.

PD 03-JUL-1997.

PF 23-DEC-1996; 96WO-GB03221.

PR 21-DEC-1995; 95GB-0026178.

PA (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.

XX Doran T, Ford J, Hermon-Taylor J, Loughlin M, Millar D;
PI Sumar N, Tizard M;

XX WPI: 1997-351061/32.

DR N-PSDB; AAT74470.

PT New isolated pathogenecity island from mycobacteria - used to
develop products for detection, diagnosis, prevention and treatment
of mycobacteria infections

PS Claim 1; Page 51; 62pp; English.

CC The present sequence represents a protein encoded by open reading frame
CC (ORF) F of a novel polynucleotide sequence designated "GS". GS is a
CC pathogenecity island of 8 kb of DNA comprising a core region of 5.75 kb
CC with multiple ORFs and an adjacent transmissible element of 2.5 kb. The
CC ORFs, and also the transmissible element, encode proteins which may be
CC linked to pathogenecity, such as providing receptors for cellular
CC recognition. GS was discovered and characterised using differential DNA
CC analysis technology. It is found within Mycobacterium paratuberculosis
CC and it has also been identified in Mycobacterium avium subspecies
CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
CC the intestine and Crohn's disease in humans. The protein products of the
CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
CC treating or preventing mycobacterial disease. In particular they can be
CC used as vaccines for inflammatory diseases such as Crohn's disease or
CC sarcoidosis in humans or John's disease in animals.

XX Sequence 266 AA;

Query Match 100.0%; Score 1401; DB 18; Length 266;
Best Local Similarity 100.0%; Pred. No. 7.3e-144;
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MTAPEVSIITPFNAAVTLQACISIVGQTYREVEVVLVDGSGTDRITDIANSFRELGS 60

QY 61 RLTVHSGPDDGPRYDANNRGVATGEMVFLGADDTLYEPTTLAQAVALGDMASHLY 120

DB 61 RLTVHSGPDDGPRYDANNRGVATGEMVFLGADDTLYEPTTLAQAVALGDMASHLY 120

QY 121 GDVYMRSTKSRHAGFPDRLLEFETNLCHOSIFRYRELEFDGIGPYNLRYVADWDFNIR 180

DB 121 GDVYMRSTKSRHAGFPDRLLEFETNLCHOSIFRYRELEFDGIGPYNLRYVADWDFNIR 180

QY 181 CFSNPALITRYMDVYISEYNDMTGFSMRGTDKEFRRLPMYFWAGWETCRMLAFKLD 240

DB 181 CFSNPALITRYMDVYISEYNDMTGFSMRGTDKEFRRLPMYFWAGWETCRMLAFKLD 240

QY 241 KENRRALRTLRIRYKAVSKERSAEP 266

DB 241 KENRRALRTLRIRYKAVSKERSAEP 266

RESULT 3

AAW21783 standard; Protein; 275 AA.

AAW21783;

DT 11-MAR-1998 (first entry)

DE Protein encoded by ORF F from the GS region of M. tuberculosis.

KW GS; pathogenecity island; pathogenic protein; mycobacterial disease;
cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
vaccine; inflammatory disease; sarcoidosis; John's disease; ss.

OS Mycobacterium tuberculosis.

PN WO9723624-A2.

AC AAB96313;
 XX
 DT 29-OCT-2001 (first entry)
 DE Putative glycosyltransferase, involved in cell wall biogenesis #1.
 XX
 KM Hyperthermophilic archaeon; hyperthermophilic protein.
 XX
 OS Pyrococcus abyssi.
 XX
 PN FR2792651-A1.
 PD 27-OCT-2000.
 XX
 PE 21-APR-1999; 99FR-0005034.
 XX
 PR 21-APR-1999; 99FR-0005034.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.
 XX
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 Querellou J, Weissenbach J, Saurin W, Heilig R;
 WPI; 2001-126236/14.
 DR
 XX New nucleotide sequences isolated from Pyrococcus abyssi encode
 PT proteins useful in industry -
 XX
 PS Claim 7; Pages 981-982; 1657pp; French.
 XX
 CC The present invention relates to the genomic sequence of Pyrococcus
 CC abyssi (see AAB96431 and AAB4123-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade.
 CC Note: This patent is in the same patent family as WO200065062, which
 CC contains additional sequences as shown in AAB9132-AAB99143,
 CC AAB75903-AAB75920 and AAG66436.
 CC
 XX
 SQ Sequence 298 AA;
 Query Match 14.7%; Score 205.5; DB 22; Length 298;
 Best Local Similarity 29.0%; Pred. No. 1.1e-13;
 Matches 71; Conservative 38; Mismatches 99; Indels 37; Gaps 10;
 1 MTAPEFIIIPFNAAYTLQACISIVGQYREVEVLVDGSDTRTDIANSFPELGS 60
 1 MSRIYVSIITIRANILIRALIASVLRQKDFELIYVDDASIDNTEVEVEST--EDGR 58
 61 ----RLVYHSGPDGDPYDANMRGVATGEWVFLGADDTLYEPTTLAQVAFLGDHAAS 116
 59 IRYIRLKKNSG--GPI-ARNIGIKKAGRFALDDDDMLPHRLLEVQVREMLGKEF 114
 117 HLYVGGDYVMASTKSR-----HAGPFDLRLLEFNLCHOSIYRRLEFDGIGPYMR 168
 115 GVYVGGFYVSQDRIILKRLPKHNGDI-YSHLKENFISPTLLIRRECFKALFDPFR 173
 169 YRVWADMDNFIRCFSPNALITR-----YMDVVISRY--NDMTGFSMRQ--GTQKEPRK 217
 174 LSSSDMDMDMIR-----IARYKFDYVDLIATKYVHGKQISNMKKYIPGERELIRK 226
 218 RLPMY 222
 227 HLDIM 231

AC AAM21779;
 XX
 DT 11-MAR-1998 (first entry)
 DE Protein encoded by ORF B from the GS region of M. tuberculosis.
 XX
 KM GS: pathogenesis island; pathogenic protein; mycobacterial disease;
 KM cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
 KM vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09723624-A2.
 PD 03-JUL-1997.
 XX
 PE 23-DEC-1996; 96WO-GB03221.
 XX
 PR 21-DEC-1995; 95GB-0026178.
 XX
 PA (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
 PA Doran T, Ford J, Hermon-Taylor J, Loughlin M, Miller D;
 PI Sumar N, Tizard M;
 PI WPI; 1997-351061/32.
 DR N-PSDB; AAT74473.
 DR
 XX New isolated pathogenicity island from mycobacteria - used to
 PT develop products for detection, diagnosis, prevention and treatment
 PT of mycobacteria infections
 XX
 PS Claim 1; Page 53; 62pp; English.
 XX
 CC The present sequence represents the protein encoded by open reading frame
 CC (ORF) B, from M. tuberculosis. This ORF B has been found to have homology
 CC with ORF B of a novel polynucleotide sequence designated "GS". GS is a
 CC pathogenesis island of 8 kb of DNA comprising a core region of 5.75 kb
 CC with multiple ORFs and an adjacent transmissible element of 2.5 kb. The
 CC ORFs, and also the transmissible element, encode proteins which may be
 CC linked to pathogenicity, such as providing receptors for cellular
 CC recognition. GS was discovered and characterised using differential DNA
 CC analysis technology. It is found within Mycobacterium paratuberculosis
 CC and it has also been identified in Mycobacterium avium subspecies
 CC silvaticum. These pathogenic mycobacteria cause chronic inflammation of
 CC the intestine and Crohn's disease in humans. The protein products of the
 CC ORFs of GS can be used for detecting mycobacteria or for diagnosing,
 CC treating or preventing mycobacterial disease. In particular they can be
 CC used as vaccines for inflammatory diseases such as Crohn's disease or
 CC sarcoidosis in humans or Johne's disease in animals.
 CC
 XX
 SQ Sequence 262 AA;
 Query Match 14.5%; Score 202.5; DB 18; Length 262;
 Best Local Similarity 32.2%; Pred. No. 1.9e-13;
 Matches 68; Conservative 36; Mismatches 90; Indels 17; Gaps 9;
 2 TAPVESIIIPFNAAYTLQACISIVGQYR-EVEVVLVDGSDTRTDIANSFPELGS 60
 3 SAPVSVITISFNDLDGLQRTVSVNRQRYRGRIEHLVIDGSGDDVVAALSGCEGFA- 61
 61 RLVYHSGPDGDPYDANMRGVATGEWVFLGADDTLYEPTTLAQ-VAAFLGDHAASHLV 119
 62 --YMOSEPDGGRYDANNOGIAHNSGDLWFLHSADRSRGDVVAQAQVEALSGGPPSEL- 118
 120 YG--DYVWSTKSRHAGPFDLRLLEFNLIC-HOSTFYRRELFDDGIPVNLRYRWADW 175
 119 WGFGRDLVGLDRVKGPIPFSLKFLAGQYVHQASFGSSLYAKIGTIDLFGLAADO 178
 176 DENIRCFSPNALITRYMDV--VISEYNDMTG 204
 179 EFLIRA-----ALVCEPVTIKCVLCEP-DTIG 204

XX	ABP26805
ID	ABP26805 standard; Protein: 321 AA.
AC	ABP26805;
DT	02-JUL-2002 (first entry)
DE	Streptococcus polypeptide SEQ ID NO 2786.
KX	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW	group A streptococcus; Streptococcus pyogenes; antibacterial;
KM	antiinflammatory; infection; vaccine; meningitis; gene therapy:
XX	Streptococcus agalactiae.
OS	WO200234771-A2.
XX	
XX	02-MAY-2002.
PF	29-OCT-2001; 2001MO-GB04789.
PR	27-OCT-2000; 2000GB-0026333.
PR	24-NOV-2000; 2000GB-0028727.
PR	07-MAR-2001; 2001GB-0005640.
XX	(CHIR-) CHIRON SPA.
PA	(GENO-) INST GENOMIC RES.
XX	Telford J, Maignan V, Margarit Ros YI, Grandt G, Fraser C;
PI	Tetella H,
DR	WPI: 2002-352536/38.
XX	N-PSDB; ABN67436.
PT	New Streptococcus protein for the treatment or prevention of infection
PT	or disease caused by Streptococcus bacteria, such as meningitis, and
PT	for detecting a compound that binds to the protein -
PS	Claim 1; Page 3429; 4525pp; English.
CC	The invention relates to a protein (ABP25413-ABP30895) from group B
CC	streptococcus/GBS (Streptococcus agalactiae) or group A Streptococcus/GAS
CC	(Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC	the specification. The proteins have antibacterial and anti-inflammatory
CC	activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC	antibodies that bind (I) are used in the manufacture of medicaments for
CC	the treatment or prevention of infection or disease caused by
CC	Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC	Nucleic acids encoding (I) are used to detect Streptococcus in a
CC	biological sample. (I) is used to determine whether a compound binds to
CC	(I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC	used as a vaccine or diagnostic composition. The disease caused by
CC	Streptococcus that is prevented or treated may be meningitis. Nucleic
CC	acid encoding (I) may be used to recombinantly produce (I) and may be
CC	used in gene therapy. Antibodies to (I) are used for affinity
CC	chromatography, immunoassays, and distinguishing/identifying
CC	Streptococcus proteins.
XX	
XX	Sequence 321 AA:
Query Match	14.4%; Score 201.5; DB 23; Length 321;
Best Local Similarity	30.1%; Pred. No. 3.4e-13;
Matches 52; Conservative 33; Mismatches 63; Indels 25; Gaps 6	
Y	7 SIITFTENAATLQAGLSIVGQTREVEVVVDGGSDTRTDIANSPFELGSRLVHS 66
D	: : : : : : : : :
B	5 SIIPVYNVSFLNCIESEVLQTSNLEIIIVNDGSTNSGDICD-YSEIDGRIFVH 63
Y	: : : : : : : : : : : : :
Y	67 GDDDPEPYAMNGVGVAIGEWYLFGADDTLYKEPTTLAOVAAFLGDHAASHLYTG----- 121
D	: : : : : : : : : : : : :
B	64 KNNGISISARRNGISRATCDYTYLDSDSYLVKKEAIEHWEE-SKYNSEIYLGCYEVR 122
D	: : : : : : : : : : : : :

[illegible]

QY 179 IRCSNPALIRIVDYISE--NDWIGFSNRGG--TDNEFRKLPMFWAGMETCRM 234
 Db 156 QGVPEGS---LRSLSLTISGVKYMDSVSSKRLRLRKELEK----- 195
 QY 235 LAFLKDKEN 243
 Db 196 IMFERNKEN 204

RESULT 9
 AAY54072
 ID AAY54072 standard; Protein; 316 AA.
 AC AAY54072;
 DT 27-MAR-2000 (first entry)
 DE Enzyme EPS8 which is involved in exopolysaccharide biosynthesis.
 KM Exopolysaccharide; EPS; EPS enzyme; EPS1; EPS2; EPS3; EPS4; EPS5;
 KM EPS6; EPS7; EPS8; EPS9; Streptococcus thermophilus strain Sfi139;
 KM activated D-galactose pyranose; saccharide; beta-1,4-galactosyltransferase;
 KM transporter; food; fermented milk product; yoghurt; cheese;
 KM flavour stability; organoleptic property.
 XX Streptococcus thermophilus.
 OS
 XX WO9962316-A2.
 PN 09-DEC-1999.
 PD
 XX 22-APR-1999; 99WO-EP02841.
 PF
 XX 22-APR-1998; 98EP-0201310.
 PR 22-APR-1998; 98EP-0201311.
 PR 22-APR-1998; 98EP-0201312.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 XX Stingele F, Germond JE, Lamothe G;
 PI
 XX WPI: 2000-097267/08.
 DR N-PSDB; AA245256, AAY54072, AAY54073, AAY54074.
 XX
 PT New recombinant enzymes for synthesis of exopolysaccharides,
 PT particularly in lactic acid bacteria, for improving properties of
 PT fermented milk products -
 XX
 PS Claim 3; Page 105-106; 162pp; French.
 XX
 AAY54065-74 represent enzymes involved in the biosynthesis of
 exopolysaccharides (EPS). These enzymes are designated EPS1-EPS10, and
 are encoded by open reading frames. eps1-eps10. The enzymes are isolated
 from Streptococcus thermophilus strain Sfi139. The proteins are used
 in a method for the synthesis of EPS, which includes at least one step
 of forming a bond (alpha or beta-isomer) between C-1 (carrying the
 reducing aldehyde function, of an activated D-galactose pyranose), and
 a phosphate on a lipophilic or proteinaceous primer. Stepwise synthesis
 of EPS occurs with, in each step, addition of a new sugar unit, through
 its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar
 unit, present at the end of a chain of sugar residues bonded to the
 primer. EPS1 to EPS4 are used to elongate the EPS chain, and to regulate
 EPS production, EPS5 creates new bonds between saccharides, EPS6 is used
 in the biosynthesis of EPS, EPS7 and EPS8 are beta-1,4-galactosyltransferases,
 CC EPS9 is a transporter of repetitive units, and EPS10 catalyses the
 CC conversion of a pyranose form of a beta-D-galactose to the furanose
 CC form. The EPS enzyme are used to improve properties of foods,
 CC particularly fermented milk products such as yoghurt and cheese,
 CC e.g. their organoleptic properties and flavour stability.
 CC
 XX Sequence 316 AA;
 SQ
 Query Match 13.3%; Score 186.5; DB 21; Length 316;

Best local Similarity 28.9%; Pred. No. 1.4e-11;
 Matches 59; Conservative 42; Mismatches 74; Indels 29; Gaps 9;

QY 4 PVFSIIPTFENNAVTLQACIGSVQTYREVEVVLVDGSTDRLDIANSFRPELSRLV 63
 Db 3 PLISIIIVPYNNEKYIRFCIESILAQTYNNIEVIYNDGSTDSLAVISLCSHNNIKV 62
 QY 64 VHSQPDGPPYDAMNRGVATGEMVLEFLGADDTLYEP---TTIAQVAAFLGDHAASHLYV 120
 Db 63 INQ-KNOGLSVARNIGIDAATGKXIAFVADADKI-KPDEVSSIXQA---DKTGADIVR 116
 QY 121 GDV-----VARSKSRHAGPDDRLLEFNILCH-QSIFYRRLPGICIPYVL 167
 Db 117 GSFREDENGNIPIKGVWDPFNVPNTYGTIVLDQFL-SSNISFVMSIYR---LDFINSNH 172
 QY 168 RYR---VMADMPENIRCFSPALI 188
 Db 173 RFTPGILFEDADFTIRAYIAKLIV 196

RESULT 10
 AAY43774
 ID AAY43774 standard; Protein; 316 AA.
 AC AAY43774;
 DT 11-FEB-2000 (first entry)
 DE Amino acid sequence of eps8 of Streptococcus thermophilus Sfi139.
 XX
 XX
 KM eps operon; Streptococcus thermophilus Sfi139; enzyme; eps1; eps2;
 KM eps3; eps4; eps5; eps6; eps7; eps8; eps9; eps10;
 KM exopolysaccharide biosynthesis; EPS; intersugar bond; antitumour;
 KM probiotic; foodstuff; organoleptic quality; flavour;
 KM lactic acid bacteria; acidified milk product; yoghurt; cheese.
 XX
 OS Streptococcus thermophilus.
 XX
 PN WO9954475-A2.
 PD 28-OCT-1999.
 XX
 XX 22-APR-1999; 99WO-EP03011.
 PF
 XX 22-APR-1998; 98EP-0201310.
 PR 22-APR-1998; 98EP-0201311.
 PR 22-APR-1998; 98EP-0201312.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 XX Stingele F, Germond JE, Lamothe G;
 PI
 XX WPI: 2000-013255/01.
 DR N-PSDB; AA230355, AAY43774, AAY43775, AAY43776;
 XX
 PT New recombinant enzymes for biosynthesis of exopolysaccharides having
 PT e.g. antitumour or probiotic properties or useful in fermented milk
 PT products -
 XX
 PS Claim 3; Page 106-107; 163pp; French.
 XX
 AAY43767-76 represent the proteins encoded by the eps operon of
 CC Streptococcus thermophilus Sfi139. The operon contains 10 open reading
 CC frames, and encodes enzymes (eps1, eps2, eps3, eps4, eps5, eps6, eps7,
 CC eps8, eps9 and eps10) that are involved in the biosynthesis of
 CC exopolysaccharides (EPS). The enzymes catalyse the formation of
 CC specific intersugar bonds. The enzymes catalyse a process that includes
 CC at least one step of forming a bond (in alpha or beta anomeric form)
 CC between C1, carrying the reducing aldehyde group of an activated D-Galp
 CC (galactose in pyranose form), and a phosphate group on a lipophilic or
 CC proteinaceous primer. The enzymes are used to produce EPS that have
 CC antitumour or probiotic properties or are used in foodstuffs to improve
 CC organoleptic qualities and flavour. When expressed by lactic acid

PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Maignant V, Margarit Ros YI, Grandi G, Fraser C

capsular component; antigen; regulation; chain length determination; complement-mediated opsonophagocytosis; serotype-specific detection.

QY 236 -APLKDKENRRAL 248
: : : : :
Db 283 IVYKOLKONKRRAL 296

Search completed: April 17, 2003, 17:26:56
Job time : 77 secs
